The Multi-Ethnic Investigation of Gene and Gene-Environment Interactions for Biochemical Markers associated with Cardiovascular Disease

Jaclyn Willette Ellis

A dissertation submitted to the faculty of the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Curriculum of Genetics and Molecular Biology

Chapel Hill
2013

Approved by

Leslie Lange, Ph.D.

Ethan Lange, Ph.D.

Karen Mohlke, Ph.D.

Kari North, Ph.D.

Monte Willis, M.D. Ph.D.

© 2013
Jaclyn Willette Ellis
ALL RIGHTS RESERVED

Abstract

JACLYN WILLETTE ELLIS: The Multi-Ethnic Investigation of Gene and Gene-Environment Interactions for Biochemical Markers associated with Cardiovascular Disease
(Under the direction of Leslie Ann Lange)

Cardiovascular disease (CVD) is the leading cause of illness and death worldwide with an estimated annual one million deaths occurring in the United States alone. Biomarkers of systemic inflammation, such as C-reactive protein (CRP), and metabolism, such as homocysteine, have been shown to be predictive of clinical CVD. CRP, an acute phase inflammatory protein, has been shown to have a moderate degree of heritability with estimates ranging between 25% and 40%. Previous reports have identified approximately twenty genetic loci to be associated with CRP. Homocysteine is a sulfur-containing nonessential amino acid derived from methionine and recently, evidence has elucidated a relationship between this amino acid and inflammation. Similar to CRP, elevated levels of homocysteine have been linked to cardiovascular complications, and heritability estimates are in the range of 47% to 70%. Further, while it has been well-demonstrated that environmental influences such as cigarette smoking significantly contribute to the development of CVD, it is less clear how genetic variants modify the effects of these risk factors on cardiovascular biomarkers. There are few studies that have assessed gene-environment interaction effects on CVD-related biomarkers. Finally, most genetic studies of CVD-related biomarkers have focused on populations of non-

European ancestry. Therefore, we assessed how common and less common gene variants influence CRP and homocysteine levels in a multi-ethnic setting, and investigated whether genetic variants in two glutathione-S-transferase genes, which play a detoxification role, modify the association between cigarette smoking and CRP level. We used African-American (AA), European-American (EA), and European (EU) participants from several cohort studies from the Candidate gene and Association Resource (CARe), the Women's Health Initiative (WHI), and Cooperative Health Research in the Region Augsburg (KORA) studies.

We identified novel genetic risk factors for CRP and homocysteine levels as well as evidence for the interplay between genes and environment on CRP. Overall, our results provide important insight regarding the pathophysiology of CVD and atherosclerosis as well as place emphasis on the need to conduct more genome-wide and candidate gene association studies in non-European populations.

I dedicate this work to my loving grandparents, the late Willie and Virginia Davis

Acknowledgments

The Lange laboratory has provided me with an excellent environment. I was able to learn far beyond my expectations. Not only have I learned genetics but concepts of epidemiology, biostatistics, and cardiovascular pathology. I want to thank both Drs. Leslie and Ethan Lange for their constant support throughout my time as a graduate student. They got me through an adverse time and I will be forever grateful.

I have an excellent committee composed of some of the most respected researchers in the University. I would like to thank Drs. Karen Mohlke, Kari North, and Monte Willis for agreeing to serve on my committee and guiding my research efforts during the course of my graduate studies. Karen provided her expertise on genetic concepts and helped me understand the importance of these concepts. Kari provided her expertise on the concepts of epidemiology which enhanced my understanding of how environment affects disease progression. Additionally, I took a course under her instruction where many of these concepts were reinforced. On multiple occasions, Monte has provided his profound expertise regarding cardiovascular pathology which gave me a much better understanding of the importance of the biomarkers I investigated. Moreover, Monte even served as a great life and career coach having a large part in guiding my career decisions. I will always be grateful for the role they played in shaping me into the scientist I am today.

I owe a great deal of appreciation to all of the investigators from the collaborating studies, the Candidate Gene and Association Resource, the Women's Health Initiative, and the Cooperative Health Research in the region Augsburg, who provided me with the data used for this dissertation. Finally, I would like to thank my older sister, Athena Ellis Chamblee, for opening up to me despite the circumstances; my younger brother, James Ellis III, for helping me to keep things in perspective; and my parents, James, Jr., and Jackie Ellis, for giving me life and providing me with continued support, love, and encouragement.

TABLE OF CONTENTS

LIST OF TABLES ... xi

LIST OF FIGURES ... xii

CHAPTER ... 1

I. Introduction ... 1

Cardiovascular Disease and Atherosclerosis ... 3

Biochemical markers ... 7

C-reactive protein ... 8

Homocysteine ... 13

Assessing Causality of C-reactive Protein and Homocysteine ... 19

Genetic Studies ... 22

Single nucleotide polymorphisms ... 23

Linkage Disequillibrium ... 23

Candidate gene studies ... 25

Genome-wide association studies/Large SNP panels ... 27

Meta-analysis ... 30

Utility of multi-ethnic samples in genetic studies ... 31

Controlling for population substructure 32

Gene-environment interactions 34

Cigarette smoking and cardiovascular disease 35

II. Large multi ethnic candidate gene study for C-reactive protein levels 39

Results 41

Discussion 44

Materials and Methods 49

III. Assessment of smoking effects on genotype and C-reactive protein levels 58

Results 60

Discussion 62

Materials and Methods 67

IV. Genome-wide analysis of African Americans and Homocysteine levels 76

Results 78

Discussion 81

Materials and Methods 87

V. Conclusions 103

Major Findings 103

Future Directions/Studies 105

Public Health Implications 108

APPENDIX I 110

Supplemental Methods 110

REFERNCES 116

LIST OF TABLES

Table 2.1. Study sample characteristics 54

Table 2.2. Loci associated with CRP in the African American samples 55

Table 2.3. Loci associated with CRP in a meta-analysis of combined samples 56

Table 2.4. Loci associated with CRP in the independent European American sample 57

Table 3.1. Participant characteristics by cohort/ethnicity 70

Table 3.2. Allelic frequencies of *GSTM1* and *GSTT1* polymorphisms 71

Table 3.3. Main effects for polymorphisms in *GSTM1* and *GSTT1* 72

Table 3.4. Interaction effects with log-pack years in European Americans 73

Table 3.5. Interaction effects with log-pack years in African Americans 74

Table 3.6. Main effects for association with log-pack years by genotype 75

Table 4.1. Descriptive table of participants by study 92

Table 4.2. Evidence of most signifcant genotype and Hcy associations 93

Table 4.3. Evidence of association at previously reported Hcy loci 94

LIST OF FIGURES

Figure 4.1a. GWAS results for inverse normal Hcy levels in men. 95

Figure 4.1b. GWAS results for inverse normal Hcy level in women... 96

Figure 4.1c. GWAS results for inverse normal Hcy levels in men and women............................ 97

Figure 4.2a. Quantile-quantile plot of SNP association with males... 98

Figure 4.2b. Quantile-quantile plot of SNP association with females.. 99

Figure 4.2c. Quantile-quantile plot of SNP association with males and females........................ 100

Figure 4.3a. Locus zoom plot of the *CSMD1* gene region 101

Figure 4.3b. Locus zoom plot of the *TC1(C8orf4)* gene region 102

CHAPTER 1

INTRODUCTION

CVD, primarily heart disease and stroke, is the leading killer in the US for both men and women among all racial and ethnic groups. The most recent statistics on mortality rates show that CVD accounts for approximately 33% of deaths in the US (Roger et al., 2011). Each year an estimated 785,000 Americans will have a new coronary attack and an additional 195,000 will experience a silent myocardial infarction (MI) (Roger et al., 2011). Furthermore, mortality rates are much higher in minority populations with the highest being in African Americans (AA) (Roger et al., 2011). A recent statistical report by the American Heart Association (AHA) stated that overall death due to CVD is greater for AA males and females than their European-American (EA) counterparts (Roger et al., 2011). In 2008, the rates per 100,000 individuals were the following: 287.2 for EA males, 390.4 for AA males, 200.5 for EA females, and 277.4 for AA females (Roger et al., 2011).

Inflammation is a complex defense mechanism in which leukocytes migrate from the vasculature to damaged tissues to inhibit agents that could cause tissue injury(Gabay, 2006). Research from clinical and population studies suggest that inflammation is an

important factor in the development of atherosclerosis and the triggering of MIs and strokes (Roger et al., 2011). Persistent inflammation may contribute to plaque instability, disruption, and ultimately thrombosis leading to a cardiac event (Gabay, 2006; Roger et al., 2011). Thus, measuring the inflammatory activity in the plasma of individuals may provide an alternative method of CVD risk prediction(Libby, Ridker, & Maseri, 2002).

The study of biomarkers can provide a better understanding of CVD development and progression. Because they can be assessed before a clinical event occurs, biomarker can be used to identify individuals at increased risk of experiencing a CVD event, which can facilitate targeted interventions. Because biomarkers are more proximal to a gene product than and offer less complexity than a CVD endpoint, the use of biomarkers as outcomes in studies may increase statistical power to detect associations (Brattstrom, Lindgren, Israelsson, Andersson, & Hultberg, 1994; Sing, Stengard, & Kardia, 2003).

Over the past decade, several biomarkers of inflammation have been linked to incident vascular events; however the most consistent results have been seen with C-reactive protein (CRP) (Conen & Ridker, 2007). CRP is a biomarker of systemic inflammation released by the cytokine interleukin 6 (IL-6) and plays an active role in the innate immune response (Ansar & Ghosh, 2013). Epidemiological studies have observed elevated levels of CRP in patients with hypertension and metabolic syndrome, which are both CVD risk factors. Homocysteine is a biomarker metabolically linked to lipid metabolism through the methionine cycle (Williams & Schalinske, 2010). Studies using animal models and human subjects have demonstrated that elevated levels of homocysteine induce endothelial dysfunction, which refers to the impairment of normal

homeostatic properties including inflammation (Cai & Harrison, 2000). Consequently, increased homocysteine levels have been associated with increased CVD risk.

This dissertation consists of three chapters that aim to assess the genetic influences of both CRP and homocysteine in a multi-ethnic cohort setting with a specific focus on African Americans. The first two chapters are aimed at using the IBC panel to identify genetic variants associated with CRP and to assess whether variants in two glutathione-S-transferase genes modify the effects of cigarette smoking exposure on CRP. The third chapter is aimed at performing a genome-wide association scan to identify loci that influence homocysteine levels in African Americans.

Cardiovascular Disease and Atherosclerosis

Cardiovascular disease (CVD) is a class of diseases that affect the cardiovascular system, primarily cardiac diseases, vascular diseases of the brain, and peripheral arterial diseases. CVD results in the interaction of multiple genes and environmental factors (Lusis, Fogelman, & Fonarow, 2004; Lusis, Mar, & Pajukanta, 2004; Q. Wang, 2005). Clot formation and atherosclerosis remain the hallmarks in the pathology of CVD (Prandoni et al., 2003). Despite pharmacological approaches to lower cholesterol levels, CVD continues to be the principal cause of death in the US, Europe, and much of Asia (Braunwald, 1997; Breslow, 1997).

Inflammation plays a critical role in the pathogenesis of atherosclerosis. The process of atherogenesis has been considered to consist largely of the accumulation of lipids within the arterial wall (Ross & Glomset, 1976a). Atherosclerotic lesions result from a highly specific cellular and molecular response with a major inflammatory

component. The lesions occur primarily in large and medium-sized elastic and muscular arteries that can lead to ischemia of the heart, brain, or extremities, and eventually resulting in infarction. The lesions can be present throughout a person's lifetime and the earliest type of lesion, called the fatty-streak (consisting of monocyte-derived macrophages and T lymphocytes), can be found in individuals as early as infancy (Napoli et al., 1997).

Endothelial dysfunction may be caused by elevated and modified low density lipoprotein (LDL) cholesterol, and free radicals caused by cigarette smoke, hypertension, diabetes mellitus, genetic alterations, or elevated plasma homocysteine concentrations (Ross & Glomset, 1973, 1976b). The result from injury to the endothelium leads to compensatory responses that alter the normal homeostatic properties of the vasculature. The different forms of injury can cause an increase in the adhesion of leukocytes or platelets, or induce procoagulant properties to form vasoactive molecules, cytokines, and growth factors. If the inflammatory response does not properly clear these agents, it can continue indefinitely; thereby, stimulating migration and proliferation of smooth muscle cells that become intermixed with the area of inflammation to form intermediate lesions (Jonasson, Holm, Skalli, Bondjers, & Hansson, 1986; van der Wal, Das, Bentz van de Berg, van der Loos, & Becker, 1989). The characteristics of an atherosclerotic lesion represent a stage of chronic inflammation. Moderate inflammation causes more macrophages and lymphocytes that migrate from the blood and multiply within the lesion. Activation of these cells leads to the release of cytokines, chemokines, and growth factors, which can induce further damage and lead to necrosis, or cell death (Falk, 1996). Continued accumulation of macrophages can lead to the enlargement and

restructuring of the lesion so that it becomes covered by a fibrous cap. Eventually, the artery will no longer be able to compensate by dilation and this advanced lesion can then intrude into the lumen and alter blood flow (Ross, 1999).

LDL cholesterol is a major contributor to injury of the endothelium and this process can be modified by oxidation, glycation, aggregation, or incorporation of the immune complex (Khoo, Miller, McLoughlin, & Steinberg, 1988; Khoo, Miller, Pio, Steinberg, & Witztum, 1992; Navab et al., 1996; Steinberg, 1997). When LDL particles become trapped in an artery, they can undergo oxidation and be internalized by macrophages via scavenger receptors on the surface of endothelial cells (Griendling & Alexander, 1997; Khoo et al., 1992; Morel, Hessler, & Chisolm, 1983; Navab et al., 1996). The internalization leads to the formation of lipid peroxidases and facilitates the accumulation of cholesterol esters, which result in the formation of foam cells. The degree to which LDL is modified can vary greatly (Diaz, Frei, Vita, & Keaney, 1997; Griendling & Alexander, 1997; Han, Hajjar, Febbraio, & Nicholson, 1997), which is why the removal and sequestration of modified LDL are important parts of the initial protective role of macrophages in the inflammation response (Diaz et al., 1997; Falcone, McCaffrey, & Vergilio, 1991; Han et al., 1997). The inflammatory response itself can have a profound effect on lipoprotein movement within the artery. Specifically, mediators of inflammation such as tumor necrosis factor α (TNFα) and interleukin-1 (IL-1) increase the binding of LDL to the endothelium and smooth muscle, and increase transcription of the LDL-receptor gene (*LDLR*). Smooth muscle cells in the media of arteries, as well as lesions, are surrounded by different types of connective tissue. In the media of arteries, the matrix consists largely of type I and III fibrillar collagen, whereas

lesions of atherosclerosis contain proteoglycan intermixed with loosely scattered collagen fibrils. Collagen inhibits cell proliferation by up-regulating specific inhibitors of the cell cycle (Koyama, Raines, Bornfeldt, Roberts, & Ross, 1996). Degradation of the collagen or its migration away from inhibitors may cause smooth muscle cells to replicate. The degradation and migration of other matrix molecules that can also inhibit the cell cycle may lead to the expression of chemokines by macrophages (Assoian & Marcantonio, 1996; Mercurius & Morla, 1998; Wesley, Meng, Godin, & Galis, 1998). Thus, the matrix surrounding smooth muscle cells greatly influence the fate of these cells (Wesley et al., 1998).

Monocyte, the precursor of macrophages in all tissues, is present at every phase of atherogeneis. Monocyte-derived macrophages are scavenging and antigen-presenting cells that secrete cytokines, chemokines, growth-regulating molecules, and other hydrolytic enzymes. Inflammatory cytokines can activate macrophages making them susceptible to apoptosis. If apoptosis occurs within the arteries, macrophages may become involved in necrotic cores which are characteristic of advanced atherosclerotic lesions (Ross, 1986).

Platelet adhesion and thrombosis are also abundant in the initiation and generation of atherosclerosis in humans and animals (Ross, 1993). Platelets can adhere to dysfunctional endothelium, exposed collagen, and macrophages. These platelets can accumulate on the walls of arteries and recruit additional platelets into a developing thrombus. Plaque rupture and thrombosis are prominent complications of advanced lesions that can result in coronary syndromes or myocardial infarction (Davies, 1990;

Falk, 1996; Ross, 1993). Moreover, plaque rupture and thrombosis may be responsible for as many as 50% of cases of acute coronary syndrome and MI (Falk, 1996).

Biochemical markers

A substantial proportion of cardiovascular events occur in individuals who exhibit none of the classical risk factors, such as age, gender, smoking, and obesity(Hackam & Anand, 2003). This critical observation has led to a collective interest in identifying biomarkers that might improve the global risk prediction of CVD (Badimon & Vilahur, 2012). Biomarkers can serve as predictors of both therapeutic and adverse clinical effects. In adverse clinical situations, when a patient is presented with severe chest pains, biomarkers of CVD can be used to differentiate whether the individual is suffering from an acute myocardial infarction or unstable angina (Vasan, 2006). In clinical trials, these same biomarkers can also be measured to determine the efficacy of the drug being tested (Vasan, 2006). Most importantly, biomarkers have also accelerated the understanding of CVD pathology (Medford, Dagi, Rosenson, & Offermann, 2013). The extensive study of cholesterols has revealed their role as key contributors to the initiation and progression of atherosclerosis. Epidemiological and clinical studies of LDL-C and HDL-C have elucidated its role in the mechanisms underlying the development atherosclerotic plaque. Exposure to elevated LDL-C levels in the endothelium decrease nitric oxide (NO) availability, which can lead to platelet aggregation, dysregulation of vascular tone, and smooth muscle cell proliferation. In contrast, elevated levels of HDL-C can protect the endothelium from such occurrences (Badimon & Vilahur, 2012).

There are multiple mechanisms underlying a complex interplay between cellular

components and inflammatory oxidative factors that contribute to the development of CVD. In addition to LDL-C, several other blood biomarkers have been established as contributing to the development of atherosclerotic plaque, such as elevated levels of homocysteine, CRP and fibrinogen (P.M. Ridker, Rifai, Rose, Buring, & Cook, 2002). Measuring these makers in the plasma of individuals may provide a greater understanding of CVD development and progression. Because biomarkers are more proximal to the gene product, the use of these as surrogate endpoints may increase statistical power to detect association(Ray, Le, Riou, & Houle, 2010).

C-reactive protein

CRP is a pentameric acute phase protein found in the blood in response to inflammation. The CRP molecule is composed of five identical non-glycosolated polypeptide subunits each containing 206 amino acid residues. The promoters are non-covalently associated in an angular configuration giving the molecule its pentameric structure (Pietila, Harmoinen, Jokiniitty, & Pasternack, 1996). The protein is mostly regulated by the cytokine interleukin-6 (IL-6), synthesized primarily by liver hepatocytes, and secreted in plasma (Morley & Kushner, 1982). Because of its ability to recognize pathogens by binding to phosphocholine (PC) resulting in the activation of the complement system, CRP presents itself as an important constituent of innate immunity (Volanakis, 2001). The innate immune system discriminates self (attack of foreign molecules) from nonself (attack of its own immunity) using a restricted number of pattern recognition receptors (PRR) that recognize pathogen associated molecular patterns (McLaughlin et al., 2002). A study of the evolution of the human immune system reveals the development of

sophisticated defense mechanisms highly specific for the invasion of pathogens (Garlanda, Bottazzi, Bastone, & Mantovani, 2005). The evolutionary conservation of mammalian CRP sequence and specificity binding over a diverse range of species (birds, amphibians, marine teleosts, and sweet water fishes) suggests that it has an important biological role. In fact, CRP is one of the first proteins to be described as a systemic biomarker for inflammation making it an important clinical biomarker (Garlanda et al., 2005). It was initially reported that only the liver can produce CRP; however, recent studies suggest that CRP may also be produced by smooth muscle cells and macrophages, found within atherosclerotic lesions linking it to CVD (Pepys & Hirschfield, 2003). There is also evidence that CRP can directly bind modified LDL cholesterol and activate the complement system, induce adhesion molecule expression, promote monocyte activation, and induce endothelial activation (Libby et al., 2002).

Circulating plasma CRP in the blood is considered normal when levels are lower than 10mg/L (Clyne & Olshaker, 1999; Das et al., 2003). In response to infection or tissue injury, CRP is reported to display marked increases in serum concentration (Morley & Kushner, 1982). For severe bacterial infections and burns, levels tend to be >200 mg/L and, for viral infection tend to be in the range of 10-40 mg/L. In response to trauma necrosis of tissue, aging, and most forms of inflammation, levels can increase about 1,000-fold 48 hours after an acute event (Clyne & Olshaker, 1999). Research has focused on the use of high-sensitivity assays to measure CRP, which allow for precision in the very low ranges to detect increased CVD risk. CRP level categories of < 1mg/L, 1-3 mg/L, and >3 mg/L are used to denote low, intermediate and high risk vascular groups, respectively (Yeh & Willerson, 2003). Thus, the quantitative estimation of CRP level in

serum is in relatively widespread use as a clinical risk predictor of CVD (Haverkate, Thompson, Pyke, Gallimore, & Pepys, 1997; Thompson, Kienast, Pyke, Haverkate, & van de Loo, 1995).

Over 20 cohort studies have demonstrated that CRP is an independent predictor of major cardiovascular events. Some studies showed that CRP was the strongest risk determinant other than age and other studies showed its magnitude of effect to be almost as great as that of blood pressure, and dyslipidemia (P.M. Ridker et al., 2002). The Women's Health Study reported that elevated CRP levels were associated with the highest rate of vascular events and FHS demonstrated that CRP enhanced CVD risk prediction. The Reynolds risk score is a measure used to predict an individual's risk of having a major cardiac event (P.M. Ridker, Buring, Rifai, & Cook, 2007). After adding information on family history of premature MI (event before age 60) and CRP to the Reynolds risk score calculation, some individuals who were initially considered to be at 10-20% disease risk were reclassified at 40-50% risk. In contrast, there is also a great deal of controversy surrounding the use of CRP as clinical biomarker. Since CRP is elevated in inflammation, the presence of any infection can limit its utility as a reliable predictor of CVD risk (Koenig, Lowel, Baumert, & Meisinger, 2004; P.M. Ridker et al., 2007). Thus, high sensitivity assays have been developed to accurately detect low concentrations of CRP to predict a healthy individual's risk of CVD (Koenig et al., 1999),.

Despite highly consistent findings demonstrating an association between CRP and CVD, it is not yet clear whether there is any direct causal role of CRP in CVD

development. Investigators of the Reykjavik Study, which included approximately 19,000 Icelandic elderly men and women, demonstrated that while the predictive value of hsCRP was statistically significant ($p<0.0001$), the magnitude of this effect was modest ($OR<1$) (Danesh et al., 2004a). The Framingham Heart Study using approximately 4500 European American men and women reported only moderate risk association between CRP and coronary heart disease ($p\sim0.09$) (Wilson et al., 2005) without the use of the Reynolds risk score. In contrast, a study using approximately 5200 elderly men and women ($\geq$ 65 years) participants from the Cardiovascular Health Study (CHS) demonstrated that CRP was an independent risk factor for ischemic stroke after ten years of follow-up where preexisting stroke was not present in the population ($P<0.05$) (Cao et al., 2003). Another study performed by CHS using African Americans (AA) and European American (EA) participants identified an association with *CRP* SNPs and CVD mortality in EAs ($P=0.004$) and myocardial infarction in AAs ($P=0.005$) (Lange et al., 2006). Even after adjustment for plasma CRP levels, which is likely influenced by *CRP* genotype, associations were only modestly attenuated but still remained significant ($P< 0.05$) (Lange et al., 2006).

CRP levels have been shown to vary across ethnic populations. Studies have reported AAs to have higher CRP levels than their white counterparts and women to have higher CRP levels than men with the highest levels observed in AA women (Khera et al., 2005). A study that examined the relationship between ethnic background and levels of inflammation showed that AAs had higher levels of CRP and IL6 than EAs; however, when adjusting for socioeconomic and health variables, the differences were attenuated for CRP but levels of IL-6 still remained higher in AAs. The elevated levels of IL-6

observed in AAs may in part explain the increased disease risk in this population and furthermore may account for the health disparities between these ethnic groups (Paalani, Lee, Haddad, & Tonstad, 2011).

Observational studies have shown CRP levels to be affected by factors such as age (Koenig et al., 1999), sex (Slade, Offenbacher, Beck, Heiss, & Pankow, 2000), smoking (Harris et al., 1999), diabetes, and atherosclerotic CVD (Mendall, Patel, Ballam, Strachan, & Northfield, 1996; R.P. Tracy et al., 1997); however, recurring evidence shows that CRP levels are also affected by genetic factors. Notably, the heritability of CRP levels is estimated to be 25% to 40% also suggesting that genetic variation is another major determinant of CRP levels (Austin et al., 2004; Dehghan et al., 2011; Fox et al., 2008; Greenfield et al., 2004; Xu & Whitmer, 2006). GWA studies have identified genetic loci to be associated with CRP. The Women's Genome Health Study (WGHS) observed six loci known to associate with CRP, which include the leptin-receptor gene (*LEPR*), *CRP* gene, interleukin-receptor gene (*IL6R*), glucokinase regulatory protein (*GCKR*), hepatic transcription factor gene (*HNF1A*), and the apolipoprotein-E (*APOE*) (P.M. Ridker et al., 2008). A recent meta-analysis of approximately 80,000 participants of European ancestry from the Cohorts for Heart and Aging Research in Genetic Epidemiology (CHARGE) consortium confirmed the association of these previously identified loci and observed significant evidence for 11 novel genes, including *NLRP3*, *HNF4A*, *RORA*, *IRF1*, and *IL1F10* (Dehghan et al., 2011). A recent GWA study of 837 individuals of African ancestry reported genome-wide significant ($P<5x10^{-8}$) associations for several variants in the CRP gene and nominal ($P<0.05$) associations for *LEPR, IL6R, GCKR, HNF1A*, and *APOE* (Doumatey et al., 2012), suggesting that these loci influence

CRP level across populations. Specific *CRP* variants associated with CRP, however, have differed between AAs and EAs. The most associated SNP in several AA studies has been rs3093058, which is only polymorphic in individuals of African descent (Carlson et al., 2005; Crawford et al., 2006), while the most associated SNP is EAs is commonly rs3091244 (Danenberg et al., 2003; Dehghan et al., 2011; Edberg et al., 2008).

Homocysteine

Homocysteine (Hcy) is an amino acid generated metabolically by the S-adenosylmethionine (SAM)-dependent transmethylation pathway (Williams & Schalinske, 2010). These reactions occur in most tissue cells; however the liver is the primary site for transmethylation and ensuing production of Hcy (Williams & Schalinske, 2010). The initial step for the metabolic production of Hcy is the activation of methionine to SAM by adenosine triphosphate (ATP)-dependent action of methionine adenosyltransferases. The outcome of all SAM-dependent transmethylation is the generation of S-adenosylhomocysteine (SAH), which is then hydrolyzed to adenosine and Hcy by SAH hydrolase (I. K. Kim, Zhang, Chiang, & Cantoni, 1983). Once Hcy has been produced, it can undergo two possible metabolic fates: remethylation back to methionine or irreversible catabolism by the transsulfuration pathway(Z. Li & Vance, 2008). Two tissue-specific Hcy remethylation pathways are known to exist: 1) a folate-dependent reaction that utilizes 5-methyltetrahydrofolate as a substrate and the action of the B_{12}-dependent enzyme, methionine synthase (MS); and 2) a folate-independent route catalyzed by betaine-Hcy S-methyltransferase (BHMT) where betaine, an oxidation product of choline, serves as the methyl donor (McKeever, Weir, Molloy, & Scott, 1991).

For both reactions, methionine serves as an acceptor of methyl groups for the maintenance of essential SAM-dependent transmethylation reactions (McKeever et al., 1991). Transsulfuration is an alternative yet essential route in the catabolism of Hcy that is initiated by the action of cystathionine β-syntase (CBS), a B_6-dependent enzyme to form cystathionine via the condensation of Hcy and serine. Cystathionine is further metabolized by ý-cystathionine to cysteine, an essential amino acid in protein synthesis and the synthesis of glutathione (Mudd & Poole, 1975). Consequently, Hcy balance is dependent on numerous reactions that result in its production; therefore, any potential disturbers of this balance could result in serious complications (Williams & Schalinske, 2010).

For folate dependent remethylation of Hcy by the B_{12}-dependent enzyme MS, most literature has focused on the polymorphic expression of the enzyme 5,10-methylenetetrahydrofolate reductase (MTHFR). For humans, a reduction in the enzymatic activity of MTHFR has been associated with hyperhomocysteinemia and vascular complications in humans (Frosst et al., 1995; Jacques et al., 1996; Nelen et al., 1998). For all cases of reduced MTHFR function, the hyperhomocysteinemia phenotype can be attributed to the compromised ability of the folate/B_{12}- dependent pathway to adequately remethylate Hcy (Beagle et al., 2005; Caudill et al., 2009; Vaughn et al., 2004). For folate-independent Hcy remethylation using betaine as a methyl donor, BHMT has been shown to be a key regulator of Hcy homeostasis. Although to date, a knock out mouse model has not been developed to understand the relationship between BHMT expression and Hcy, a pharmacological inhibitor of BHMT (S-(-carboxybutyl-$_{DL}$-Hcy)) caused an increase in total plasma Hcy levels by sevenfold (Collinsova, Strakova,

Jiracek, & Garrow, 2006). For the irreversible catabolism of Hcy, previous literature has focused primarily on CBS, the B_6-dependent enzyme. In humans, it was observed that a lack of CBS expression was linked to homocystinuria, characterized by an increased excretion of Hcy in the urine (Mudd, Finkelstein, Irreverre, & Laster, 1965). Using a mouse model, Wang, et al. demonstrated that the induction of hepatic and renal CBS lowered serum Hcy concentrations (L. Wang et al., 2004). Hyperhomocysteinemia has been linked to many disease states in both mice and humans; however, the leading emphasis has been its recognition as an independent risk factor for cardiovascular disease.

Evidence has demonstrated that Hcy has atherogenic and prothrombotic properties. There are multiple mechanisms on how Hcy may induce vascular injury or endothelial dysfunction. It can promote leukocyte recruitment by upregulating the expression and secretion of chemokines such as chemoattractant protein-1 and interleuking-8 (Poddar, Sivasubramanian, DiBello, Robinson, & Jacobsen, 2001). The thiol metabolite of Hcy can combine with LDL-cholesterol to produce aggregates that are susceptible to uptake by macrophages in the arterial intima; the ensuing foams cells can lead to the development of atherosclerotic plaques (McCully, 1996). Free radicals formed during the oxidation of reduced Hcy may cause direct injury to endothelial cells (Mansoor, Bergmark, Svardal, Lonning, & Ueland, 1995; Starkebaum & Harlan, 1986) and result in marked platelet accumulation (McCully & Carvalho, 1987; Stamler et al., 1993). Prolonged exposure to Hcy in endothelial cells reduces the activity of the enzyme, asymmetric dimethylarginine (ADMA), responsible for the degradation of nitric oxide synthase (eNOS), which may contribute to impaired endothelium vasodilation (Kanani et

al., 1999; Tawakol, Omland, Gerhard, Wu, & Creager, 1997; Woo et al., 1997). The association between elevated Hcy plasma levels and venous thrombosis, coronary heart disease (CHD), and atherosclerosis has been demonstrated repeatedly (Cattaneo, 1999; Wald, Law, & Morris, 2002; Welch & Loscalzo, 1998). Hyperhomocysteinemia accounted for 13% of population attributable risk (PAR) for MI (Whincup et al., 1999) and for 10% of PAR for coronary artery disease (CAD) (Boushey, Beresford, Omenn, & Motulsky, 1995). High plasma concentrations of Hcy were said to be associated with advanced atherosclerotic lesions in patients with defects in enzymes MTHFR and CBS. It was later observed that patients with these defects can develop severe atherosclerosis starting as early as childhood, and many have their first MI by the age of 20 years (Husemoen, Thomsen, Fenger, Jorgensen, & Jorgensen, 2003; Kluijtmans et al., 2003).

The regulation of plasma Hcy is a complex phenotype because it is influenced by multiple factors. The main environmental determinants are gender, age, renal function, and vitamin intake (folate, B12, and B6). However, it has been demonstrated that genetic determinants play a role in the variation of Hcy levels across populations (Gaustadnes, Rudiger, Rasmussen, & Ingerslev, 2000; Kang, 1995; Mezzano et al., 1999; Smulders, de Man, Stehouwer, & Slaats, 1998), with the heritable portion of variation in Hcy across individuals estimated to range between 47%-70%. The first genetic determinant identified for Hcy in the general population was a C to T transition at position 677, in the methylenetetrahydrofolate reductase (*MTHFR)* gene, which encodes for the MTHFR enzyme. Carriers of the T allele produce a less-efficient form of the MTHFR enzyme so homozygotes of the T allele exhibit a moderate increase in Hcy levels (Gudnason et al., 1998)(54-Suoto). This polymorphism is estimated to explain 6% of the variance in

plasma Hcy levels (Husemoen et al., 2003) with a higher proportion explained in males versus females (7.3% and 5.1% respectively) (Lange et al., 2010). Several studies have investigated whether genetic associations of this variant can be generalized to multiple populations, the clinical utility of this variant, and its prevalence across multiple populations

The Coronary Artery Risk in Young Adults (CARDIA) study investigated the role of *MTHFR* and other genes involved in Hcy metabolism to determine its polygenic influence pre and post folic acid fortification era in the US. In 1996, the United States Food and Drug Administration (FDA) mandated that all enriched grain products be fortified with folic acid to reduce the risk of neural tube defects in newborns. As a result, there was a decrease in the prevalence of hyperhomocysteinemia (Honein, Paulozzi, Mathews, Erickson, & Wong, 2001). Between years 0 and 7(pre-fortification era), the 677C>T variant was associated with Hcy in N=997 EA ($p< 0.001$) and N=692 AA ($p<0.003$). However, post fortification era (year 15), significance was only observed in EA ($p=0.004$). Notably, another variant in the MS gene (A2756G) was associated with Hcy in AA both before fortification ($p<0.03$) and after fortification ($p=0.007$) (Tsai et al., 2009). The same group then assessed the clinical utility of the 677C>T variant in predicting hyperhomocysteinemia across both EA and AA. Because the prevalence of the minor allele T in the 677C>T variant is so low, meaningful assessment could only be carried out in European Americans. Sensitivity of *MTHFR* 677C>T genotyping in predicting hyperhomocysteinemia was >60% pre folic acid fortification and 30% post folic acid fortification (Tsai et al., 2009). Conversely, the specificity for 677C>T was

much higher (87-88%) and this percentage remained unchanged even after folic acid fortification (Tsai et al., 2009).

As observed in the CARDIA study, the prevalence of *MTHFR* gene polymorphisms varies across different ethnic groups. Esfahani, et al. determined the prevalence of two common polymorphisms in *MTHFR* (677C>T and 1298A>C) across women of four different ethnic groups: Hispanic women of Mexican descent, Asian women, African American, and European American women. The frequency of the homozygous TT genotype was the highest in Mexican women and lowest in AA women (Esfahani, Cogger, & Caudill, 2003). The frequency of the minor allele (C) in the 1298A>C variant was highest in EA women. They also assessed the frequency of combined genotypes for both variants. Mexican and EA women had the highest frequency of combined heterozygosity (heterozygous for both mutations (17.6% and 15.1% respectively) and AA and Asian women had the lowest (6.3% and 3.8% respectively) (Esfahani et al., 2003). An international study was conducted to determine the prevalence of the *MTHFR* 677C>T variant in 7130 newborns across 16 areas in the Americas, Europe, Russia, China, and Australia (Wilcken et al., 2003) based on geographic location and ethnicity. Based on geographic location, the prevalence of the *MTHFR* TT genotype was 10-12% in Spain, France, and Hungary; however in Finland, Helenski and northern Netherlands, its prevalence was between 4-6%. In areas of southern Europe (Campania and Sicily), the frequency was much higher (26% and 20% respectively). In New South Wales and Canada, the frequencies were 7.5% to 6% respectively. In the Americas, the frequency was highest in Mexico (32%) In the US (Atlanta, GA), homozygosity of the T allele was more common in newborns of Hispanic

origin, intermediate among newborns of European ancestry in Atlanta (11%), and lowest among newborns of African ancestry. It is evident that there is an impact of geographical and ethnic variation favoring the presence of this gene variant (Wilcken et al., 2003). Furthermore, this also indicates that genes that may be ethnic specific could have a stronger influence on Hcy levels.

Several GWA studies have also identified other loci to affect Hcy levels. The WGHS performed a GWA scan to identify novel variants that may influence Hcy levels in approximately 14,000 EA women. They identified four new loci associated with plasma Hcy levels: *carbamoyl phosphate synthetase I* (*CPS1)* (rs7422339 at 2q34), *methylmalonyl-CoA mutase MUT* (rs4267943 at 6p12.3), *nicotinamide adenine dinucleotide phosphate oxidase NOX4* (rs11018628 at 11q14.3), and *dipeptidase 1* (*DPEP1*) (rs1126464 at 16q24.3) (Pare et al., 2009). A GWA study of Hcylevels in Filipino women and their offspring from the Cebu Longitudinal Health and Nutrition Survey (CLHNS) reported evidence of sex-dependent effects for *CPS1* variant (rs7422339). In the pooled offspring sample, there was only evidence for association in the female subset but no evidence in males. Furthermore, the direction of effect in the female offspring was in the same effect as the mothers (Lange et al., 2010) as well as for the WGHS. To date, no GWA studies with plasma Hcy levels have been performed in a primarily AA dataset.

Assessing Causality of C-reactive Protein and Homocysteine

Although several studies have demonstrated associations of both CRP and Hcy with CVD, their roles as causal factors are still under debate. Epidemiologic studies have tried

to investigate the cause-and-effect relationship between biomarkers and CVD outcomes using an approach called Mendelian randomization (Lawlor, Harbord, Sterne, Timpson, & Davey, 2008). This approach is based on Mendel's second law, which states that alleles of different genes assort independently during gamete formation (Wensley et al., 2011). Mendelian randomization uses common genetic polymorphisms that are known to influence the risk factor of interest as proxies of the risk factors that are less susceptible to confounding by behavioral or environmental factors (Sheehan, Didelez, Burton, & Tobin, 2008). There are three key assumptions for Mendelian randomization studies:

1. The genetic variant must be unrelated to known confounding factors for the relationship between the intermediate trait and the outcome.
2. The genetic variant is associated with the outcome and this can be accurately quantified.
3. For known exposure status and known confounders, the genetic variant is independent of the outcome (Sheehan, Meng, & Didelez, 2011).

Therefore, in the case of CVD related biomarkers, this method should reduce confounding factors if the variants related to the exposure are unrelated to any other disease markers.

One study using 47,000 Danish individuals from the Copenhagen General Population Study used Mendelian randomization to assess causality between CRP and atrial fibrillation. An increase in disease risk was seen in individuals with elevated CRP levels within the general population but this correlation was not repeated in the portion of measured CRP estimated to be genetically elevated (based on genotype of previously

established CRP variants), suggesting that circulating CRP does not play a causal role in atrial fibrillation (Marott et al., 2010). A study conducted by the CRP Coronary Heart Disease Genetics Collaboration (CCGC) examined the role of CRP in relation to coronary heart disease (CHD). Similarly, they did not identify a causal role for CRP in CHD. However, this does not rule out a causal role for inflammation in CVD development as one or more of the assumptions made for Mendelian randomization analysis may not be valid.

Mendelian randomization studies have also been conducted to investigate the causality between Hcy and *MTHFR* genotype. A study consisting of participants from the National Health and Nutrition Examination Survey (NHANES) used the *MTHFR* C677T variant in Mendelian randomization to examine whether Hcy has a causal role in CVD mortality and all-cause mortality (Yang et al., 2012). As expected, the *MTHFR* C677T genotype was associated with higher Hcy concentrations as well as lower serum folate concentrations. Specifically, individuals with the TT genotype had a 2.2 µmol/L higher Hcy concentration on average than individuals with the CC genotype (Yang et al., 2012). A folate concentration > 3.4 ng/mL was generally associated with a lower risk of all-cause and CVD mortality; however, Hcy concentrations were not significantly associated with either of the clinical endpoints (Yang et al., 2012). Interestingly, the *MTHFR* TT genotype was associated with a lower risk of CVD mortality (HR=0.69, p=0.026) compared to the *MTHFR* CC genotype (control) even after stratified analyses by sex, ethnicity, use of folic acid supplements, serum folate deficiency status, and CVD status at baseline (Yang et al., 2012). A study using 12,239 postmenopausal women from the Netherlands, without folic acid fortification, also implemented Mendelian

randomization to assess for a causal relationship between the *MTHFR* C677T variant and CVD mortality. The incident rates for CVD were highest for *MTHFR* CC wild types and lowest for TT homozygotes even after adjustment for age. The rate ratios for CVD mortality were 0.7 for women with the CT genotype and 0.6 for women with the TT genotype compared with the CC genotype (0.8), however there was no evidence for a causal relationship (P>0.05) (Roest et al., 2001). In contrast, a study performed on another Dutch cohort of 666 elderly men and women (≥ 85 years) found the *MTHFR* 677C>T genotype to be associated with CVD mortality (P=0.03). However, gender stratified analysis revealed that the association was attenuated in women (P=0.51) and became more significant in men (P=0.01) (Heijmans et al., 1999). These contradictory findings could suggest that the influence of the *MTHFR* genotype is dependent on differences in folate status between men and women.

Genetic Studies

The primary goal of human complex disease genetics is to identify genetic risk factors for common diseases such as schizophrenia, type II diabetes, and CVD (Bush & Moore, 2012). There are many different technologies, study designs, and analytical tools for identifying genetic risk factors, yet ultimately, the goal is to use these genetic risk factors to make predictions about disease risk and determine the biological underpinnings of disease susceptibility to develop new and more effective therapeutics (Bush & Moore, 2012).

Single nucleotide polymorphisms

The unit of genetic variation is the single nucleotide polymorphism (SNP). SNPs are single base-pair changes in the DNA sequence that occur with high frequency in the human genome (Genomes Project et al., 2010). SNPs are typically used as markers of a genomic region where most of these markers have minimal impact on biological systems. However, SNPs can have functional consequences some of which include an amino acid change, changes to mRNA transcript stability, and transcription factor binding affinity (Griffith et al., 2008). SNPs typically have two alleles meaning there are two commonly occurring base-pair possibilities for a particular SNP location. Mendelian diseases are usually caused by a single, rare genetic mutation that induces a detrimental change to protein function leading to the diseased state, with high penetrance. For common complex disease, however, SNPs that are identified are often not themselves functional and are merely tagging variants with usually very modest effects, and there are often many such variants spread across the genome.

Linkage Disequillibrium

Linkage disequilibrium (LD) is a property of SNPs on a contiguous stretch of a genomic sequence that describes the degree to which a SNP allele is inherited or correlated with an allele of a different SNP. The concept is closely related to chromosomal linkage, which relates how two markers on a chromosome remain physically joined through multiple family generations. Generally, recombination events within a family from generation to generation break apart chromosomal segments (Bush & Moore, 2012). LD decreases through increasing generations of random mating, which continue to break apart

contiguous chromosomal segments. Under random mating and no selection or migration, eventually all of the alleles in a population will reach linkage equilibrium or become independent of each other.

Patterns of LD can help identify founder effects (Bush & Moore, 2012). The rate of LD decay is dependent on multiple factors including population size, number of founding chromosomes, and the number of generations for which the population has existed (Bush & Moore, 2012). As such, different subpopulations, or ethnic groups, have varying degrees of LD. Populations of African descent have smaller regions of LD due to the accumulation of more recombination events. European and Asian descent populations relatively recently descended from sub-African populations; thus, these groups on average have larger regions of LD due to less recombination (Bush & Moore, 2012).

Two commonly used measures of LD are D' and r^2 (Devlin & Risch, 1995; International HapMap, 2005); yet for the purposes of genetic analysis, LD is generally reported in terms of r^2, a statistical measure of correlation (Fallin & Schork, 2000). SNPs that are selected specifically to capture the variation at nearby sites in the genome are called tag SNPs because alleles for these SNPs tag the surrounding stretch of LD. Based on analysis of data from the HapMap project, >80% of commonly occurring SNPs in European descent populations can be captured using a subset of approximately 500,000 to one million SNPs across the genome (M. Li, Li, & Guan, 2008). SNP-trait associations can result from either direct or indirect association (Hirschhorn & Daly, 2005). In a direct association, the SNP influencing a biological system that leads to the

phenotype is directly genotyped. In an indirect association, the influential SNP is not directly genotyped but rather a surrogate for the causal SNP is studied (Hirschhorn & Daly, 2005). Due to these two possible outcomes, a significant SNP association should not be assumed as causal (Bush & Moore, 2012).

Candidate gene studies

Candidate gene studies investigate genes that are selected for their potential role in the etiology of the disease or trait based on known biology or previous studies that have identified the gene in relation to the disease or phenotype of interest. Thus, a candidate-gene study takes advantage of both increased statistical efficiency of the association analysis of complex diseases and the biological understanding of the phenotype and the genes under consideration (Tabor, Risch, & Myers, 2002). The selection of candidate genes is very similar to the identification and ranking of risk factors in an epidemiological study (Terwilliger & Goring, 2000). The first step is to examine published studies of the phenotype of interest. Evidence can be evaluated for the involvement of specific genes, which may influence the phenotype of pathological disease models or expression studies that can provide information about the tissues and cells that are involved in the disease (Perou et al., 2000; Scherf et al., 2000; Welsh et al., 2001). Subsequently, the range of possible hypotheses, the size of the study population, and the magnitude of the effect of the genes is also evaluated (Long & Langley, 1999).

The next step is to prioritize SNPs based on their functionality. Theoretically, it is desirable to only study polymorphisms that are likely to affect the function of a protein or its expression. However, information about the functionality of all variants is often

difficult to obtain or not established. In addition to considering the function of polymorphisms, it is also important to consider their frequency in the population being tested for association. The statistical power to detect a significant association depends on the effect size of the association and the frequency of the allele of interest (Lalouel & Rohrwasser, 2002; Long & Langley, 1999; N. Risch & Merikangas, 1996). SNPs with very low allele frequencies would need to have very large relative risks associated with them to be detected in a candidate gene study; therefore, SNPs with frequencies of at least 5% are generally more likely to be useful (N. J. Risch, 2000). Because variants with severe functional consequences tend to be much less frequent in the population, more common variants with less severe consequences may have to be given a higher priority in candidate-gene study. Lastly, another important consideration in selecting SNPs is whether there is significant LD in the candidate gene within the study population. Determining LD in a small pilot sample can help to optimize SNP selection and provide information about possible or analysis (Lewontin, 1988). If several SNPs in a gene or gene region are in high LD, it is possible to infer some of the genotypes based on the genotypes of a subset of SNPs. Therefore, the total number of genotypes assayed can be reduced by selecting SNPs that are most likely to be functional from a set of SNPs that are in significant LD (Drysdale et al., 2000; Subrahmanyan, Eberle, Clark, Kruglyak, & Nickerson, 2001).

A limitation of candidate gene studies is that the current knowledge regarding disease pathologies as well as gene functions may not be sufficient to allow for optimal selection of genes. However, candidate gene studies can be a cost-efficient and powerful approach, particularly in the following cases: to follow up genes identified in one

population in additional populations; to fine-map genes where associations with likely non-causal variants have previously been reported; and to evaluate gene-by-environment interactions in settings that are biologically plausible.

Genome-wide association studies/Large SNP panels

The idea that common diseases have a different underlying genetic architecture than rare disorders, coupled with the discovery of several susceptibility variants for common diseases with a high minor allele frequency, led to the development of the common disease/ common variant (CD/CV) hypothesis (Reich & Lander, 2001). The hypothesis simply states that common disorders are likely influenced by multiple genetic variants that are each common in the population. Genome-wide association studies (GWAS), which consist of evaluating panels of tag SNPs across the entire genome for association with the trait of interest, are now widely used to identify susceptibility loci and operate under the CD/CV hypothesis. However, there are several key inferences made in the use of this particular hypothesis.

1. If the common genetic variants influence disease, the effect size for any one variant must be small relative to that found for rare disorders.
2. Second, if common alleles have small genetic effects, or low penetrance, but common disorders show heritability, then multiple common alleles must influence disease susceptibility.

Essentially, if the allele of a single SNP incurs only a small degree of risk, that SNP only explains a small proportion of the total variance due to genetic factors. Per se, the total genetic risk due to common genetic variation must be spread across multiple genetic

factors (Bush & Moore, 2012). As such, common variants cannot have high penetrance (Bush & Moore, 2012).

A systematic approach is needed to examine much of the common variation in the human genome. First, the location and density of commonly occurring SNPs is needed to identify the genomic regions. Second, population-specific differences in genetic variation must be cataloged so that studies of phenotypes in different populations can be conducted under the proper study design. Finally, correlations among common genetic variants must be determined so that genetic studies do not collect redundant information (Bush & Moore, 2012). The International HapMap Project was designed to carry out such an approach. The project uses a variety of sequencing techniques to discover and catalog SNPs in the following populations: European, Yoruba of African origin, Han Chinese from Beijing, China, and Japanese from Tokyo, Japan (International HapMap, 2005; Ritchie et al., 2010).

Exploitation of LD patterns allows for the use of a method called imputation where GWAS datasets are imputed to generate results for a common set of SNPs without directly genotyping (Y. Li, Willer, Sanna, & Abecasis, 2009). The concept is similar in principle to haplotype phase algorithms, where the contiguous set of alleles lying on a specific chromosome is estimated. Genotype imputation extends this methodology to populations. First, a collection of shared haplotypes within the study sample is computed to estimate haplotype frequencies among previously genotyped SNPs. Next, phased haplotypes from the study sample are compared to reference haplotypes from a panel of much more dense SNPs, such as HapMap or 1000 Genomes.

The study sample haplotypes may match multiple reference haplotypes so the surrounding genotypes are given a score of probability instead of actually assigning the imputed SNP a particular allele (i.e. A). The probabilities for each allele at each SNP are estimated, for example *(0.85 A, 0.12 C, 0.03 T*), based on estimated haplotype frequencies (Biernacka et al., 2009; Browning & Browning, 2009; Guan & Stephens, 2008; B. N. Howie, P. Donnelly, & J. Marchini, 2009).

Often the effects identified in an initial GWAS suffer from the winner's curse, where the observed effect is likely stronger in the discovery GWAS sample as compared to the general population. Therefore, the gold standard for the validation of any genetic study is replication in an additional sample. There are several criteria, which have been outlined by a working group from the National Human Genome Research Institute (NHGRI), that establish a positive replication (Studies et al., 2007). First, replication studies should have sufficient sample size to detect the effect of the susceptibility allele. Second, they should be conducted in an independent dataset drawn from the same population as GWAS, in an attempt to confirm the effect of the GWAS target population. Replication of a significant result in an additional population (other than the target) is sometimes called generalization meaning the genetic effects are relevant across multiple populations. Third, identical phenotype criteria should be used in both the GWAS and replication study. Lastly, a similar effect should be seen in the replication set from the SNP, or a SNP in high LD with the GWAS-identified SNP (Bush & Moore, 2012).

The results of multiple GWAS studies can be pooled to perform a meta-analysis. Meta-analysis techniques were developed to examine and refine significance and effect

size estimates from multiple studies. A fundamental principle in meta-analysis is that all studies are testing the same hypothesis; therefore, the general design of each study should be similar. Quality control procedures that determine which SNPs are included; covariate adjustments; and the measurement of clinical covariates and phenotypes should all be standardized. Additionally, the sample sets across all studies should be independent. The most critical standard is to verify that all studies report results relative to a common effect or coding allele. If one study reports reference allele A, whereas, the other reports reference allele B, the meta-analysis result may not be significant because the effects will nullify each other (Sanna et al., 2008; Willer et al., 2008). Because it is rare to find multiple studies that match all this criteria, heterogeneity is often statistically quantified to determine the degree to which the studies differ (Huedo-Medina, Sanchez-Meca, Marin-Martinez, & Botella, 2006).

Meta-analysis

Recognizing the need and benefits of data sharing, GWAS investigators have formed various networks or consortia to share data on the same disease or related disorders (Dehghan et al., 2011; Ioannidis, 2008; Pankow et al., 2001; Sobrin et al., 2011). The general perception is that a mega-analysis (raw, individual data) is statistically more efficient because it utilizes much more detailed information. However, a caveat for obtaining such data is that it is costly and time-consuming. Therefore, another preferred method for pooling data is conducting a meta-analysis (Lin & Zeng, 2010). Any genetic effect produced by the accumulation of individual raw data in a mega-analysis or the pooling of data in a meta-analysis have approximately the same variance (Lin & Zeng, 2010).

A meta-analysis is a method of systematically combining pertinent and qualitative and quantitative study data from several selected studies to develop a single conclusion that has greater statistical power (Bahekar, Singh, Saha, Molnar, & Arora, 2007). This conclusion is statistically stronger than the analysis of any single study due to increased numbers of subjects, greater diversity among subjects, or accumulated effects and results (Bahekar et al., 2007). The advantages of conducting a meta-analysis are greater ability to gain statistical power; establishing a confirmatory data analysis; greater ability to extrapolate to the general population affected; and this analysis is considered to be an evidence-based resource (Ageno, Becattini, Brighton, Selby, & Kamphuisen, 2008). However, one of the major concerns for perform such analysis is publication bias of literature results. This bias can be reduced by systematically selecting studies, which use a standardized phenotype and similar statistical methodologies (i.e. models, covariate adjustments) (Bush & Moore, 2012). Additionally, the creation of consortiums can also reduce this bias. In a consortium setting, the same covariate adjustments can be made within each study and the covariate-adjusted estimates of genetic effects can then be combined through a meta-analysis (Lin & Zeng, 2010).

Utility of multi-ethnic samples in genetic studies

It has been demonstrated that the use of non-European samples can improve the power to detect susceptibility loci (N.A. Rosenberg et al., 2010). Pulit, et al. showed that minor allele frequencies differences across populations can lead to dramatically different statistical power to detect associations for those populations (Pulit, Voight, & de Bakker, 2010).

Because allele frequencies differ across populations, the incorporation of multiple ethnic samples can help elucidate ethnically distinct loci and/or variants. For example, the SNP (rs2283228) in the *KCNQ1* locus, associated with Type 2 diabetes, has a minor allele frequency of ~40% in East Asian populations but only ~5% in European populations (McCarthy, 2008; Unoki et al., 2008; Yasuda et al., 2008). Similar to allele frequencies, patterns of linkage disequilibrium (LD) vary across populations; thus the study of multiple populations can facilitate fine-mapping that could uncover a true causal polymorphism. A study by Shriner, et al., using participants of African and Asian ancestry, localized a region in chromosome 1 consisting of loci previously identified in EAs to be associated with human height (Shriner et al., 2009). It should be noted, however, that this approach makes the assumption that the same causal variant is underlying the association across ethnicities, which may not be accurate in some cases.

Controlling for population substructure

Stratification in genetic studies is often attributed to differences in population genetic structure and substructure (Tian, Gregersen, & Seldin, 2008). If these ancestry differences are unaccounted for, it can lead to false positive findings. Even when studies are restricted to a single continental origin, many false-positive results may still be observed due to subtle differences in ethnic make-up (Clayton et al., 2005; Marchini, Cardon, Phillips, & Donnelly, 2004; A. L. Price et al., 2006). To address this issue, statistical methods can be applied to discern and correct for these differences.

Generally, three approaches are used to correct for population substructure. These methods are structured association tests; principal component analyses (PCA), and multidimensional scaling (MDS). Structured association depends on applying information from model-based or distance-based clustering algorithms. The model-based STRUCTURE program (Falush, Stephens, & Pritchard, 2007; Pritchard, Stephens, & Donnelly, 2000) can be used to perform tests conditional on group (or ethnic) membership. This membership is determined using a Bayesian clustering algorithm that fits the data to the number of specified cluster groups. This method tests the null hypothesis, which is that there is no dependence of allele frequencies on phenotypes within each group; however, the limitation of this tests is that it requires an estimation on the number groups being tested and is computationally intensive (Tian, Gregersen, et al., 2008). When Euclidean distance is used, classical metric MDS is the same as PCA. Both PCA and MDS can infer a continuous axis of genetic variation independent of assigning individuals to various subpopulations or groups.

Both approaches reduce high-dimensional data to smaller numbers of dimensions that intuitively group patterns together based on the observed data (Tian, Plenge, et al., 2008). The program EIGENSTRAT (A.L. Price et al., 2006) calculates ancestry-adjusted genotypes and phenotypes using the continuous axis of variation from PCA to compute the association statistic (Tian, Gregersen, et al., 2008); however, the number of principal components (PCs) that need to be examined will vary across datasets. Therefore, to determine the number of PCs that should be considered, a genomic control parameter should be implemented (Tian, Plenge, et al., 2008).

Substantial differences in allele frequencies within different continental populations have been reported (N. A. Rosenberg et al., 2002). In general, the number of SNPs showing large allelic frequency differences between major continental populations are an order of magnitude greater than those seen within continental populations. Thus, the largest source of type 1 errors will be caused by differences in the distribution of ancestry from major continental populations. Self-identification of ancestry usually reduces this problem; however, the presence of admixture can lead to bias results (N. A. Rosenberg et al., 2005). PCA and MDS will distinguish an individual with substantial admixture; however the use of ancestry informative markers (AIMs) can be used to verify self-identification of ethnicity (Tian, Gregersen, et al., 2008).

Gene-environment interactions

The concept that the etiology of most common diseases involves not only discrete genetic and environmental factors but the interaction between the two is increasingly being accepted (Garrod, 1996). In the context of genetics and epidemiology, the study of gene-environment interactions is extremely useful. If we estimate only the separate contributions of genes and environment to a disease, ignoring their interactions, we will incorrectly estimate the proportion of the disease that is explained by genes, environment, and their joint effect (Hunter, 2005). Ultimately, understanding gene-environment interactions might further allow the use of individualized preventative measures, in addition to offering personalized therapeutics post diagnosis (Hunter, 2005).

Screening a larger number of potential interactions in datasets with a large number of genotypes and many variables of exposure greatly increases the chances of

identifying a false-positive (T. R. Rebbeck, Spitz, & Wu, 2004). In addition, most studies do not individually have sufficient sample sizes for detection modest interactions (T. R. Rebbeck et al., 2004). Restricting the search for gene-environment interactions to biologically plausible candidate genes and utilizing meta-analysis across studies addresses these issues.

Cigarette smoking and cardiovascular disease

Smoking causes 140,000 premature deaths from cardiovascular disease annually in the US representing about 30% of all smoking-related deaths (Burns, 2003). While the specific mechanisms are not understood, cigarette smoking acts synergistically with other cardiovascular risk factors to increase CVD morbidity and mortality (Burns, 2003; Unverdorben, von Holt, & Winkelmann, 2009). Epidemiological studies strongly support the idea that cigarette smoking in both men and women increases the incidence of MI and fatal CAD. Current smoking is associated with an eighty percent increase in CAD risk and second-hand tobacco exposure, which is approximately $1/100^{th}$ of that of an active smoker, is associated with a 30% increase in CAD risk (Glantz & Parmley, 1991; Law, Morris, & Wald, 1997). Although evidence has suggested a link between cigarette smoke exposure and CVD, the exact components of cigarette smoke and the mechanisms responsible for the association have not been clearly elucidated (Ambrose & Barua, 2004).

Vasomotor dysfunction, inflammation, and modification of lipids precede the clinical manifestations of atherosclerosis (Clarkson, Weingand, Kaplan, & Adams, 1987; Ross, 1999). Endothelial dysfunction results in the impairment of vasodilatory function,

which is one of the earliest manifestations of atherosclerotic changes in the blood vessel. The endothelium produces its vascular effects via the release several small molecules including nitric oxide (NO), prostacyclin, tissue plasminogen activator (tPA), and plasminogen activator inhibitor-1 (PAI-1) (Benowitz, 2003). Cigarette smoking may alter these chemicals, which could result in vasoconstriction (Shen, Rattan, Sultana, & Kalra, 1996). Smokers without atherosclerosis have coronary vasoconstrictor effects in the endothelium that in the presence of normal function would produce vasodilation; these effects are greater in the presence of hypercholesterolemia (Adams, Jessup, & Celermajer, 1997; Weber, Erl, Weber, & Weber, 1996). Several studies indicate that cigarette smoking causes an approximate 20%-25% increase in the peripheral blood leukocyte count (C. J. Smith & Fischer, 2001) and is associated with an increased level of multiple inflammatory markers including CRP, IL-6, and TNFα in both male and female smokers (Bermudez, Rifai, Buring, Manson, & Ridker, 2002; Tappia, Troughton, Langley-Evans, & Grimble, 1995; R. P. Tracy et al., 1997). Elevations of proinflammatory cytokines increase leukocyte recruitment to the site of endothelial injury (Bermudez et al., 2002; Mazzone et al., 2001). Cigarette smoking also causes activation of proatherogenic molecules leading to an alteration in intercellular interactions. Cigarette smoking exposure was associated with 70% to 90% increase in the adherence between human monocytes and human umbilical vein endothelial cells (HUVECs) due to increased expression of adhesion molecules on the surface of both monocytes and HUVECs (Kalra et al., 1994).

Although cigarette smoking is extraordinarily hazardous, many smokers do not appear to suffer injury from their addiction. Fifty-percent of life-long smokers die

prematurely from smoking-related diseases, whereas, 50% do not (Benowitz, Jacob, Jones, & Rosenberg, 1982). Moreover, some smokers develop severe CVD, which may result in an early death but some appear to be resistant to CVD despite a history of heavy smoking (J. Wang et al., 2010). These sources of variability may be due to genetic modifications of risk which may both enhance susceptibility and protect against disease.

Most of the literature on genetic influences of smoking-induced disease in humans has largely been focused on a family of genes called the glutathione-S-transferases (GSTs). GSTs have been one of the primary classes of genes evaluated as they are involved in the detoxification and/or activation of some chemicals in cigarette smoke(T.R. Rebbeck, 1997). GSTs catalyze the conjugation of glutathione, an important antioxidant, to a variety of electrophilic compounds and peroxidases (Anderson, 1998). This family of genes is divided into six classes where the mu and theta classes seem to contribute primarily to the interindividual differences in pathology or clinical response (Townsend).

The mu (*GSTM1*) and theta (*GSTT1*) class of *GST* have a null phenotype (often designated as *GSTM*0* and *GSTT1*0*) whereby individuals do not express the catalytically active enzyme. The lack of enzymatic activity is associated with an increase risk to a variety of cancers and heart diseases (Townsend, Tew, & Tapiero, 2003). The *GSTM1*0* allele is observed in ~50% of the European American population and is associated with an increased risk of lung, colon, and pulmonary asbestosis (C. M. Smith et al., 1994; Strange, Jones, & Fryer, 2000). In contrast, the *GSTT1*0* allele varies between ethnic groups but is most prevalent in Chinese populations (65%) and lowest in

Mexican American populations (9%) (Nelson et al., 1995). The *GSTT1* null phenotype is associated with an increase in tumors of the head, neck, and oral cavity (Chenevix-Trench, Young, Coggan, & Board, 1995; Strange, Lear, & Fryer, 1998). Thus, GST-related genes are ideal candidates for assessing as potential modifiers of the effect that smoking behavior has on inflammatory CVD biomarkers, such as CRP.

CHAPTER 2

LARGE MULTI ETHNIC CANDIDATE GENE STUDY FOR C-REACTIVE PROTEIN LEVELS

C-reactive protein (CRP) is a pentameric acute-phase protein that is a hallmark of systemic inflammation.(P.M. Ridker, 2010) Vascular inflammation is thought to play a role in the development and progression of atherosclerosis, ultimately leading to plaque rupture and cardiovascular disease (CVD) events such as myocardial infarction.(Robbie & Libby, 2001) Associations between CRP and CVD outcomes have been remarkably consistent despite varying study designs, target populations, and case classification methods.(Pankow et al., 2001) Observational studies have shown increased levels of CRP to be present in individuals with factors such as older age, obesity, (Koenig et al., 1999) female sex,(Slade et al., 2000) smoking,(Harris et al., 1999) diabetes, atherosclerotic CVD,(Mendall et al., 1996; R.P. Tracy et al., 1997) sleep curtailment (van Leeuwen et al., 2009) and sleep apnea.(Larkin et al., 2005) Familial studies have reported heritability estimates for CRP levels of 35-40%, (Schnabel et al., 2009) supporting that genetic factors are likely to influence the variation of CRP levels in response to CVD risk factors.(Pankow et al., 2001)

Multiple candidate gene and genome-wide association studies (GWAS) have been performed for CRP resulting in several reported associated loci. These associations include genes known to be involved in the regulation of inflammatory and metabolic

pathways, some of which were not previously known to directly influence CRP levels. The Women's Genome Health Study identified single nucleotide polymorphisms (SNPs) associated with CRP in the leptin-receptor gene (*LEPR*), glucokinase regulatory protein (*GCKR*), and hepatic transcription factor gene (*HNF1A*).(A.P. Reiner et al., 2008) Recently, a GWAS meta-analysis of participants of European ancestry confirmed association of previously identified loci with CRP and introduced 11 novel loci, including *NLRP3*, *HNF4A*, *RORA*, *IRF1*, and *IL1F10*.(Dehghan et al., 2011) To our knowledge, only two published GWASs for CRP in individuals of African ancestry have been conducted (Doumatey et al., 2012; A. P. Reiner et al., 2012), and the first was based on a relatively small cohort of individuals.(Doumatey et al., 2012) This earlier study identified several variants in the *CRP* gene that were associated with CRP, but no other loci were statistically significant. (Doumatey et al., 2012) The latter study, which included 8,280 African American (AA) women from the Women's Health Initiative (WHI) study, also identified a number of variants associated with CRP in the *CRP* gene as well as significant evidence for associations in or near *IL1F10/IL1RN*, *TREM2*, *HNF1A* and *TOMM40/APOE*.(A. P. Reiner et al., 2012)

We sought to extend what is known regarding the genetic underpinnings of CRP by performing a multi-ethnic meta-analysis, including individuals of both African and European ancestry genotyped across a densely covered gene-based array. Participants for the primary analyses came from eight community-based cohorts from the Candidate Gene and Association Resource (CARe) consortium (AAs and European Americans [EAs]), WHI (EAs) and the Cooperative Health Research in the Region of Augsburg (KORA) study (Europeans). All participants had available genotype data from the

ITMAT Broad-CARe (IBC) Chip, a custom 50,000 SNP gene-centric array having dense coverage of over 2,000 candidate genes within CVD related pathways. Additional AA participants from WHI study were used as a follow-up sample for interesting findings.

Results

Characteristics of the 7,584 AA and 29,975 European-ancestry study participants from CARe, WHI and KORA can be found in Table 2.1.

African Americans

Four loci reached study-wide significance ($p<2.2x10^{-6}$) in AAs (Table 2.2), including three loci reported in previous GWASs for CRP in individuals of European descent (*CRP*, *IL6R*, *APOE*). The fourth significant result, at rs3211938 in the gene encoding the cluster differentiation 36-membrane protein (*CD36*; $p= 1.4x10^{-6}$), has not been reported previously. A locus zoom plot of the *CD36* region is presented in Figure 1, demonstrating the focused association at rs3211938. We further tested this SNP using imputed exome sequencing data in an independent sample of 8,041 African Americans women from WHI. The association was confirmed, with the minor allele of *CD36* rs3211938 associated with 0.128 (+/- 0.030) lower CRP levels ($p=1.8x10^{-5}$). Together, in a combined meta-analysis of CARe and WHI African American participants, results at *CD36* rs3211938 reached genome-wide significance ($p =1.5x10^{-10}$). Of note, *CD36* rs3211938 was not studied in the recent WHI GWAS. (A. P. Reiner et al., 2012) There was no evidence for any heterogeneity in the results across the AA cohorts for any of the four significant loci (Table 2.2).

Out of the five loci previously reported to be associated with CRP in AA women by Reiner, et al., we had good proxies for SNPs at genes *CRP* and *TOMM40* but no available satisfactory proxies for the remaining gene regions (*HNF1A*, *ILF10/IL1RN*, and *TREM2*) (A. P. Reiner et al., 2012).

Our top result occurred at *CRP* rs3093058 ($p=4.2x10^{-71}$) (Table 2.2), which is in strong linkage disequilibrium (LD) with our best available proxy for rs16827466 ($r^2=0.89$) (based on 1000 Genomes data in subjects of African ancestry (YRI)), the top variant reported by Reiner et al.(A. P. Reiner et al., 2012) Our most significant result at *APOE* was for SNP, rs769450 ($p=2.0x10^{-6}$), which is in modest LD ($r^2=0.28$) with the top SNP, rs1160985 in nearby *TOMM40*, reported by Reiner et al. However, our best proxy for SNP rs1160985 (rs405509, $r^2=0.66$) demonstrated nominal evidence for association with CRP ($p=1.2x10^{-4}$). We had poor proxies available for the top SNPs reported for the other three significant loci in Reiner et al. Our best proxy for *IL1F10/IL1RN* rs6734238, (rs17042795, $r^2=0.28$) demonstrated a trend towards association (p=0.068), as did our best proxy for *HNF1A* rs7979473 (rs1169293, $r^2=0.33$, p=0.052). We did not find any evidence for an association at our best proxy (rs6933067, $r^2=0.39$, p=0.14) for the reported novel *TREM2* SNP rs7748513 association.

Combined Race Meta-Analysis

We observed significant signals ($p<2.2x10^{-6}$) at 13 loci, including seven loci widely reported to be associated with circulating CRP levels (*CRP, TOMM40/APOE/APOC1, HNF1A, LEPR, GCKR, IL6R, BAZ1B/ BCL7B and IL1RN*) and

two (*NLRP3* and *HNF4A*) recently reported as significant by the Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) consortium (Table 2.3). We identified a CRP-associated locus, the aryl hydrocarbon receptor nuclear translocator like gene (*ARNTL*), that was reported by CHARGE to be "suggestive" (i.e. nominally associated with $p<1.0x10^{-5}$ with CRP levels), and two novel loci, the gene encoding the ribosomal protein S6 kinase (*RPS6KB1)* and *CD36*.

The direction and size of the effects were largely consistent between AAs and subjects of European descent for most of these 13 loci (Table 2.3). In some cases, the frequency of the minor allele was considerably different between the populations, which could largely explain the absence of significant evidence for association in one population or the other. One SNP, rs2075650 in *TOMM40*, had highly discrepant results between AAs and subjects of European descent. There was no evidence for any association for the variant in AAs; in fact there was a near trend for an effect in the opposite direction as was observed in subjects of European descent. Interestingly, both AAs and subjects of European descent had evidence for association at nearby SNP rs769450 (AAs see Table 2; Europeans, beta (SE) =0.033(0.009), $p=1.0x10^{-4}$). The two variants are in weak LD (estimated r^2 in YRI = 0.088, r^2 in CEU = 0.13) in populations of African and European ancestry.

Analyses of European Americans not included in CHARGE consortium

A subset of the current study samples was also included in the CHARGE report, which included ~66,000 subjects of European descent with CRP and genotype measurements.(Dehghan et al., 2011) Genotype data used in the CHARGE report were

obtained across a variety of genome-wide marker platforms; genotype imputation was used to probabilistically infer missing genotype data for SNPs in the HapMap database. The current study, which includes ~38,000 samples, has ~22,000 overlapping European-ancestry samples with the CHARGE report. Notably, no African Americans were included in the CHARGE report.

Focusing on the CHARGE EA sample (n~16,000) not previously identified in CHARGE, we identified significant evidence for six loci (*CRP*,*TOMM40/APOE/APOC1*, *HNF1A*, *LEPR*, *GCKR, IL6R*) widely recognized to be associated with CRP (Table 2.4). We found evidence supporting two new reported associations in the CHARGE report (*HNF4A* rs1800961 p=5.4x10^{-4}; *NLRP3* rs12239046 p=0.011), providing, to our knowledge, the first reported confirmation for these findings. We also found indirect evidence, through a proxy SNP, supporting another novel CHARGE finding, rs13233571 in the *BAZ1B/BCL7B* gene cluster. We found nominal evidence at *BAZ1B* rs714052 (p=7.6x10^{-3}), which is in strong LD with the reported CHARGE SNP (r^2=0.93 in CEU HapMap samples). Additionally, we found indirect evidence, through proxy SNP rs6486121 (p=0.0025), supporting an association at rs6486122 in *ARNTL* reported to be nominally significant in CHARGE. Interestingly, in this non-overlap EA sample, there was no evidence of association at *RPS6KB1* rs1292034 (p=0.41), which we identified as a novel locus in our combined meta-analysis.

Discussion

We performed a dense candidate gene-based scan of approximately 50,000 SNPs covering approximately 2,000 gene regions in a combined bi-racial sample of ~32,000

individuals. We observed significant evidence for 7 loci widely reported to be associated with CRP, which include: *CRP, TOMM40/APOE/APOC1, HNF1A, LEPR, IL6R, GCKR, IL1RN.* We also found IBC array-wide significant evidence for association *at HNF4A, NLRP3,* and *BAZ1B/BCL7B*, loci that were reported to be significantly associated with CRP in the recent CHARGE report. Analyses based on an independent subset of samples not included in the CHARGE report provide supporting evidence for these associations. We also found significant evidence for association at *ARNTL*, a locus reported as suggestive in the CHARGE report. Finally, we identified two novel loci, *CD36* and *RPS6KB1,* which have not been previously reported to be significantly associated with CRP. The *CD36* association at rs3211938 is specific to African Americans. A meta-analysis at rs3211938 including an additional sample of 8,041 African American women from WHI resulted in a genome-wide significant association.

The *CD36* gene encodes a cellular receptor that facilitates fatty acid uptake and the utilization of key metabolic tissues.(Coburn et al., 2000; Tanaka et al., 2001) Individuals with mutations in *CD36* have a defective FA uptake which could result in a poor metabolic profile and elevated serum lipid levels. (Miyaoka et al., 2001; Tanaka et al., 2001) Patients with elevated lipid levels (i.e. LDL cholesterol) are more likely to develop atherosclerosis, which most broadly defines cardiovascular disease risk. Various scavenger receptors have been recognized for their role in mediating the uptake of oxidized LDL (ox-LDL) leading to the formation of foam cells which is a precursor step in the development of atherosclerotic lesions.(Goyal et al., 2012) *CD36* is a scavenger receptor involved in the uptake of these oxidized lipids and thus plays a role in the formation of atheroma.(Goyal et al., 2012). Interestingly, an *in vitro* experiment showed

that the addition of CRP to ox-LDL in a cell culture system stimulated foam cell formation suggesting that CRP may have an active role in enhancing foam cell formation. (Ji, Wu, Potempa, Qiu, & Zhao, 2006) Together, these findings suggest that CRP and *CD36* may have a cooperative role in atherogenesis.

The influence of natural selection, linked to malaria susceptibility, has resulted in the high genetic variation of *CD36* in populations of African descent (Aitman et al., 2000; Omi et al., 2003); as a result, *CD36* mutations are more commonly observed in African populations versus populations of primarily European descent. (Love-Gregory et al., 2008). The link to malaria susceptibility results in the gene's role as a receptor for *Plasmodium falciparum* infected erythrocytes, which is found in malaria patients. Variants in this gene have also been reported be associated with metabolic syndrome, (Love-Gregory et al., 2008) HDL cholesterol levels (Love-Gregory et al., 2011; Musunuru et al., 2012), and abnormal serum FA. (Handberg, Levin, Hojlund, & Beck-Nielsen, 2006; Silverstein, 2009; Sun et al., 2010) A study using African American participants from HyperGEN demonstrated that *CD36* variants account for ~3.4% of inter-individual HDL variability in the study population.(Coon et al., 2001) The nonsense variant rs3211938 identified in our analysis encodes for a truncated form of the CD36 protein and causes an amino acid change from threonine to glycine. This variant has only been identified in populations of African descent, having thought to arise due to positive selective pressure, and thus its predictive impact on CD36 expression would not apply to other populations.(Fry et al., 2009; Sabeti et al., 2006) In a recent study by Love-Gregory, et al., the minor allele (G) for rs3211938 was associated with increased HDL and reduced CD36 protein expression on monocytes. (Love-Gregory et al., 2011) The

same coding allele (G) for rs3211938 was associated with lower CRP levels in our analysis, suggesting that *CD36* SNPs may elucidate a biological link between CRP and HDL levels. A GWAS study on platelet count and mean platelet volume of 16,388 African American individuals identified novel associations at two intronic SNPs at the *CD36* gene (Qayyum et al., 2012). Previous evidence has demonstrated that one of the variants (rs17154155) is associated with platelet function as well as platelet expression of CD36 (Ghosh et al., 2011; Jones et al., 2009), and interestingly, rs17154155 happens to also be in LD with the rs3211938 variant associated with CRP levels in our analysis ($r^{2}=$ 0.271). Further analysis on less common *CD36* variants in African Americans from the National Heart Lung and Blood Institute Exome Sequencing Project found rs3211938 to be associated with lower platelet count in this population (P. L. Auer et al., 2012). Taken together these results also suggest the potential for *CD36* variants to be used as a predictive tool for CVD risk in African Americans (Love-Gregory et al., 2011).

A recent study from the CARe consortium, using the same AA participants included in this study, also found strong evidence for an association between HDL levels and *CD36* rs3211938. (Musunuru et al., 2012) HDL and CRP levels are modestly correlated across these cohorts (e.g. Spearman's correlation for ARIC= -0.050 and JHS= -0.037). Inclusion of HDL as a covariate in the linear models for the CARe samples only modestly impacted the association between CRP and rs3211938 ($p = 1.4x10^{-6}$ for the model unadjusted for HDL vs. $p = 5.4x10^{-6}$ for the HDL-adjusted model), suggesting that the *CD36* association with CRP levels is largely independent of HDL.

We also found evidence at *ARNTL*, also termed *BMAL1*, which is a core component of the circadian clock and a vital element of the central circadian pacemaker.

(Honma et al., 1998) This locus had suggestive evidence for association in the CHARGE report. Molecular circadian clocks exist in peripheral tissues and coordinate gene transcription involved in a wide range of metabolic processes including gluconeogenesis, lipolysis, adipogenesis, and mitochondrial oxidative phosphorylation to achieve an appropriate internal alignment of metabolic signaling as well as external alignment of cellular processes. (Buijs et al., 2006) It has been well documented that pathologic events display circadian rhythms with an increase in incidences, such as myocardial infarction and ischemia, from dawn to noon.(Maemura et al., 2001; Portman, 2001) Furthermore, experimental circadian misalignment is associated with abnormalities in blood pressure, glucose, and insulin levels (Scheer, Hilton, Mantzoros, & Shea, 2009) while shift work, a real-world model for circadian misalignment, is associated with diabetes and CVD.(Karlsson, Knutsson, & Lindahl, 2001; Knutsson, Akerstedt, Jonsson, & Orth-Gomer, 1986) Polymorphisms of *ARNTL* have been associated with age at menarche. Inactivation of *ARNTL* results in altered regulation of blood pressure, lipid metabolism, and glucose homeostasis; and these changes have been observed in hypertensive mouse models.(Bunger et al., 2000; Curtis et al., 2007; Naito et al., 2003; Rudic et al., 2004; Shimba et al., 2005) The association between *ARNTL* and CRP suggests one possible mechanism of correlation between CRP and metabolic dysregulation, thereby increasing CVD risk.

Strengths of this study include the large African-American and European-ancestry dataset with measured CRP levels the large commercial candidate-gene-based genotyping panel. The generalizability to other populations remains to be determined. SNPs identified may not be causally linked to variation in CRP concentration levels, rather they

may be in linkage disequilibrium with the causal variants. In addition, the functional mechanisms linking genetic variants to CRP concentrations remain to be determined. Finally, we had limited power to detect low frequency variants and SNPs with low effect size, especially in AAs. We acknowledge that additional genomic loci may be uncovered in larger samples and with broader coverage of genetic variation across the human genome.

Our findings provide better insight about the pathways involved in variation in CRP concentrations and provide additional data implicating the role of circadian genes in influencing cardiovascular risk. Overall, these findings are consistent with the role of metabolism and inflammatory pathway in the regulation of circulating CRP levels.

Materials and Methods

Each study was reviewed by a local ethics board and all participants consented to genetic research. Genotype and phenotype data for all study participants, with the exception of KORA subjects, are available through the NCBI dbGaP resource (www.ncbi.nlm.nih.gov/gap).

Study samples

CARe

The CARe (Candidate Gene Association Resource) consortium consists of nine studies. The purpose of the consortium was to bring together deeply-phenotyped prospective cohort studies to increase power for genetic association scans of CVD and other disorders. (Musunuru et al., 2012) Cohorts included in these analyses of CRP levels are:

Atherosclerosis Risk in Communities (ARIC) (n=12,617), Coronary Artery Risk in Young Adults (CARDIA) (n=2,738), Cleveland Family Study (CFS) (n=1,349), the Cardiovascular Health Study (CHS) (n=4,704), Framingham Heart Study (FHS) (n=7,556), Jackson Heart Study (JHS) (n=2,036), and Multi-Ethnic Study of Atherosclerosis (MESA) (n=3,910). Further details of the participating CARe studies are reported in the **Supplemental Methods.**

Women's Health Initiative (WHI)

WHI is one of the largest (n=161,808) studies of women's health ever undertaken in the U.S. ("Design of the Women's Health Initiative clinical trial and observational study. The Women's Health Initiative Study Group," 1998) There are two major components of WHI: (1) a Clinical Trial (CT) that enrolled and randomized 68,132 women ages 50 – 79 years into at least one of three placebo-control clinical trials (hormone therapy, dietary modification, and calcium/vitamin D); and (2) an Observational Study (OS) that enrolled 93,676 women of the same age range into a parallel prospective cohort study. ("Design of the Women's Health Initiative clinical trial and observational study. The Women's Health Initiative Study Group," 1998) A diverse population was recruited from 1993-1998 at 40 clinical centers across the U.S.

KORA

The MONICA/KORA (MONItoring of trends and determinants in Cardiovascular disease/ Cooperative Health Research in the Region of Augsburg) Augsburg study is a series of population-based surveys conducted in the region of Augsburg in Southern Germany.(Lowel et al., 2005) The data for the present study was drawn from a subcohort

randomly selected by sex and survey from surveys S1 to S2 conducted between 1984 and 1990 (KORA S12) (Thorand et al., 2005) and from survey F3 conducted 2004/05 (KORA F3). After exclusion of subjects with no phenotype or IBC genotype data, with individual call rate < 0.90 and signs of relatedness (DST > 0.95), the study populations consisted of 1,075 participants for KORA S12 and 1,800 participants for KORA F3. All participants were from European ancestry.

Genotyping

CARe

The IBC SNP array is described in detail in Keating et al. (Keating et al., 2008) The IBC SNP array includes 49,320 SNPs selected across ~2,000 candidate loci for CVD. The array includes SNPs that capture patterns of genetic variation in both European- and African-descent populations. Genotyping for the CARe cohorts was performed at the Broad Institute (Cambridge, MA). Criteria for DNA sample exclusion based on genotype data included sex mismatch, discordance among duplicate samples, or sample call rate <95%. For each set of duplicates or monozygotic twins, data from the sample with the highest genotyping call rate were retained. SNPs were excluded when monomorphic, the call rate was <95%, or when significant departures from expected Hardy-Weinberg equilibrium (HWE) genotype proportions were observed ($p<10^{-5}$ in EAs). Given the genetic admixture in African Americans, there was no HWE filter used for these samples. After these exclusions were applied, data remained on 47,539 SNPs.

Women's Health Initiative

DNA was extracted by the Specimen Processing Laboratory at the Fred Hutchinson Cancer research Center (FHCRC) using specimens that were collected at the time of enrollment. Only participants with CRP measured at baseline were included in this analysis.

Genome-wide genotyping was performed at Affymetrix using the Affymetrix 6.0 array. A total of 8,421 AAs had genotype data that passed quality control. A reference sample of 761 AA NHLBI Exome Sequencing Project (ESP) participants was used for imputation of *CD36* rs3211938 into 8,041 AA WHI individuals with GWAS data using the programs MaCH 1.0.18 (Y. Li, Willer, Ding, Scheet, & Abecasis, 2010) and minimac (B.N. Howie, P. Donnelly, & J. Marchini, 2009) Additional details on the genotype imputation are given in Auer at al. (Auer, Srivastava, & Doerge, 2012)

KORA

Genotyping was done within the HumanCVD Genotyping Kit (IBC 50K Chip) as suggested by the manufacturer (Illumina, Inc., San Diego, USA). The genotyping was carried out using the CVD-Upenn_Phase1 manifest with the Beadstudio software (Genotyping module v.3.2.29, no call threshold 0.15).

Data analysis

Participants with CRP measurement over 10 mg/L were excluded from analysis as these observations would be highly influential and potentially a result of acute infection. We natural log-transformed CRP level in order to generate an approximately normal distribution of model residuals, conditional on the covariates, to meet linear model assumptions.

We assumed an additive genetic model in all tests of association. We used the linear regression model implemented in PLINK (Purcell et al., 2007b) for studies with unrelated individuals and the linear mixed effects model implemented in the program GWAF for cohorts with related individuals (M.H. Chen & Q. Yang, 2010) to test for association between log-CRP and genotype at each SNP, adjusted for covariates. All models were stratified by cohort and race. Covariate adjustment was applied for age, sex, current smoking, body mass index and the first 10 principal components calculated using the program EIGENSTRAT (A.L. Price et al., 2006) to control for potential population substructure. After obtaining cohort- and race-specific results, we performed a fixed-effects, inverse variance-weighted meta-analysis using the METAL software.(C.J. Willer, Y. Li, & G.R. Abecasis, 2010) Meta-analysis was performed separately by race and race-combined.

Based on a simulation analysis performed by Lettre and colleagues, the effective number of independent tests was calculated to be 26,482 for the African Americans and 20,544 for samples of primarily European ancestry, accounting for the linkage disequilibrium between markers on the IBC array. To maintain an overall type 1 error rate of 5%, a uniform statistical threshold of $\alpha = 2.2 \times 10^{-6}$ (0.05/25000) was thus used to declare array-wide (experiment-wide) significance.(Nalls et al., 2011)

Table 2.1. Study sample characteristics (% or mean±SD)

	ARIC		**CARDIA**		**CFS**		**CHS**		**FHS**	**JHS**	**MESA**		**WHI**	**KORA**
	EA	**AA**	**EA**	**AA**	**EA**	**AA**	**EA**	**AA**	**EA**	**AA**	**EA**	**AA**	**EA**	**EU**
N	7581	1987	1332	1126	281	371	3919	736	7556	2026	2051	1338	4389	2866
Female, %	53.9	64.1	53.2	59.2	53.0	58.0	56.3	62.5	53.8	60.6	52.0	53.7	100.0	48.9
Current Smoker, %	20.9	26.1	23.0	29.3	24.6	18.3	11.1	16.1	50.0	15.1	10.7	17.9	7.6	21.8
Diabetes, %	7.6	16.5	0.53	0.85	11.1	17.6	14.7	24.6	10.9	15.4	5.8	16.2	7.3	4.5
Age, y	54.1 ±5.7	52.9 ±5.7	25.7 ±3.3	24.4 ±3.8	44.3 ±19.3	40.4 ±18.6	72.8 ±5.6	73.0 ±5.7	48.8 ±13.7	50.0 ±12.0	62.2 ±10.1	61.7 ±9.8	68.2 ± 6.5	52.1 ±10.7
Body Mass Index, kg/m2	26.9 ±4.7	29.6 ±5.9	23.7 ±4.0	25.6 ±5.7	31.7 ±9.2	33.1 ±9.8	26.4 ±4.5	28.5 ±5.6	27.4 ±5.5	32.2 ±7.8	27.7 ±5.1	30.1 ±5.7	27.6 ± 6.8	27.2 ±4.1
CRP, mg/L	4.02 ±5.6	5.8 ±7.1	2.4 ±3.7	4.3 ±6.3	3.8 ±5.2	4.6 ±6.2	4.3 ±6.7	6.1 ±8.2	3.3 ±6.2	5.1 ±7.7	3.4 ±5.1	4.7 ±6.9	4.2 ± 6.1	2.8 ±5.2

EA=European American; AA=African American; EU=European

Table 2.2: Loci associated with CRP in the African American samples

Chr.	No. SNPs $P<2.2 \times 10^{-6}$	Most Significant	Position	Allele	Freq.*	Beta(SE)*	het p-value	p-value	Gene
1q23	13	rs3093058	157951939	T/A	0.17	0.39(0.022)	0.69	4.2×10^{-71}	*CRP*
1q21	2	rs8192284	152693594	A/C	0.14	-0.12(0.025)	0.60	2.0×10^{-6}	*IL6R*
7q21	1	rs3211938	80138385	T/G	0.08	-0.14(0.030)	0.22	1.6×10^{-6}	*CD36*
19q13	2	rs769450	50102284	G/A	0.37	0.083(0.017)	0.16	1.6×10^{-6}	*TOMM40*

* Allele frequency and beta estimate is presented for second allele.

Table 2.3. Loci associated with CRP in a meta-analysis of African Americans (AA) and subjects of European (Eur) descent.

Chr.	Most Significant	Alleles	Position	Freq. (Eur)*	Beta(SE) (Eur)*	p-value (Eur)	Freq. (AA)*	Beta(SE) (AA)*	p-value (AA)	Meta-analysis p-value	Gene
1p31	rs1805096	G/A	65874845	0.38	-0.098(0.009)	$3.3x10^{-30}$	0.45	-0.061(0.017)	$2.5x10^{-4}$	$2.6x10^{-32}$	*LEPR*
1q21	rs4129267	C/T	152692888	0.40	-0.079(0.008)	$5.2x10^{-21}$	0.13	-0.12(0.025)	$5.7x10^{-7}$	$1.2x10^{-24}$	*IL6R*
1q23	rs3091244	A/G	157951289	0.62	-0.17(0.009)	$3.5x10^{-91}$	0.45	-0.24(0.017)	$5.1x10^{-45}$	$7.8x10^{-132}$	*CRP*
1q44	rs12239046	C/T	245668218	0.39	-0.039(0.009)	$5.3x10^{-6}$	0.51	-0.064(0.017)	$1.4x10^{-4}$	$7.3x10^{-9}$	*NLRP3*
2p23	rs1260326	C/T	27584444	0.43	0.094(0.008)	$5.3x10^{-29}$	0.14	0.056(0.024)	0.019	$1.0x10^{-29}$	*GCKR*
2q13	rs4251961	T/C	113590938	0.38	0.060(0.009)	$1.4x10^{-12}$	0.18	0.081(0.022)	$2.0x10^{-4}$	$1.8x10^{-15}$	*IL1RN*
7q11	rs714052	A/G	72502805	0.12	-0.065(0.013)	$3.9x10^{-7}$	0.042	-0.056(0.044)	0.20	$1.7x10^{-7}$	*BAZ1B*
7q21	rs3211938	T/G	80138385	<0.001	-0.42(0.54)	0.44	0.085	-0.14(0.030)	$1.6x10^{-6}$	$1.4x10^{-6}$	*CD36*
11p15	rs6486121	T/C	13312606	0.37	-0.043(0.009)	$6.4x10^{-7}$	0.58	-0.037(0.017)	0.026	$1.3x10^{-7}$	*ARNTL*
12q24	rs2244608	A/G	119901371	0.34	-0.11(0.009)	$4.6x10^{-39}$	0.14	-0.076(0.024)	0.0016	$9.5x10^{-41}$	*HNF1A*
17q23	rs1292034	G/A	55344642	0.45	-0.037(0.008)	$6.0x10^{-6}$	0.82	-0.039(0.02)	0.079	$1.2x10^{-6}$	*RPS6KB1*
19q13	rs2075650	A/G	50087459	0.13	-0.18(0.012)	$2.2x10^{-47}$	0.13	0.039(0.02)	0.11	$1.9x10^{-34}$	*TOMM40*
20q13	rs1800961	C/T	42475778	0.030	-0.13(0.024)	$2.0x10^{-7}$	0.0070	-0.14(0.10)	0.18	$7.8x10^{-8}$	*HNF4A*

* Allele frequencies and beta estimates are presented for second allele.

Table 2.4. Loci associated with CRP in the independent European American replication sample *(CARDIA,CFS, MESA, WHI)*

Chr.	No. SNPs $P<2.2x10^{-6}$	Most Significant	Position	Allele	Freq.*	Beta(SE)*	het p-value	p-value	Gene
1p31	19	rs1805096	65874845	G/A	0.38	-0.14(0.017)	0.61	$2.5x10^{-18}$	*LEPR*
1q21	5	rs4129267	152692888	C/T	0.40	-0.10(0.016)	0.69	$6.1x10^{-10}$	*IL6R*
1q23	9	rs3091244	157951289	G/A	0.38	0.18(0.016)	0.61	$2.9x10^{-28}$	*CRP*
2p23	3	rs1260326	27584444	C/T	0.43	0.085(0.016)	0.35	$1.5x10^{-7}$	*GCKR*
12q24	5	rs2244608	119919810	A/G	0.31	-0.12(0.017)	0.38	$1.2x10^{-14}$	*HNF1A*
19q13	6	rs12721046	50087459	G/A	0.13	0.18(0.023)	0.36	$7.3x10^{-15}$	*APOC1*

Allele frequency and beta estimate is presented for second allele.

CHAPTER 3

ASSESSMENT OF SMOKING EFFECTS ON *GSTM1* AND *GSTT1* GENOTYPE EFFECTS AND C-REACTIVE PROTEIN LEVELS

Glutathione S-transferases (GSTs) are a family of Phase II enzymes that are responsible for the metabolic inactivation of electrophilic compounds and toxic substrates (Conklin et al., 2009), (Tirona & Pang, 1999; Townsend, Findlay, & Tew, 2005; Wu, Fang, Yang, Lupton, & Turner, 2004). GSTs are divided into three main families: cytosolic, mitochondrial, and membrane-bound. The cytosolic family is further divided into seven classes which include the mu and theta classes encoded by the genes *GSTM1* and *GSTT,1* respectively. Many epidemiological studies have focused on these two gene classes because these isoforms are polymorphically deleted (Board, 1981) and likely to contribute to inter-individual differences in response to xenobiotics and clearance of oxidative stress products (J. Wang et al., 2010). Thus, it has been suggested that these variations may determine susceptibility to various inflammatory pathologies including cardiovascular disease (CVD). Furthermore, studies have also suggested that these GST variations may alter CVD risk to a greater extent in cigarette smokers than non-smokers (Conklin & Bhatnagar, 2011).

Cigarette smoking is a common habit that affects about 23% of the adult US population (Jockel et al., 2009) being causally linked to myocardial infarction, stroke,

coronary artery disease and other forms of CVD (Yusuf et al., 2004). Because evidence has shown cigarette smoking to be associated with systemic inflammation, it has been hypothesized that inflammation may be one of the mechanisms through which smoking affects CVD (Unverdorben et al., 2009).

Increased levels of C-reactive protein (CRP), a pentameric acute-phase protein, is a trademark of low-grade systemic inflammation (P. M. Ridker & Silvertown, 2008) and a biomarker used in CVD risk prediction. The association between CRP and smoking has been established by many epidemiological studies. (Asthana et al., 2010; Benowitz, 2003; Burns, 2003; Danesh et al., 2004b) Increased CRP levels were found in former and current smokers (Lowe, Yarnell, Rumley, Bainton, & Sweetnam, 2001), and only those former smokers who had quit for more than 20 years had CRP levels comparable to their never smoking counterparts (Wannamethee et al., 2005). It has been well demonstrated that circulating levels of CRP are influenced by smoking (Z. Chen & Boreham, 2002) and genetic factors (Danesh et al., 2004b; Pepys & Hirschfield, 2003; R. P. Tracy et al., 1997); however, it is not yet clear how smoking influences genetic effects on CRP levels.

Candidate genes studies have explored whether smoking effect on CRP is modified by genetic variation. The glutathione-S-transferases (GSTs) have been one of the primary classes of genes evaluated because they are involved in the detoxification and/or activation of some chemicals in cigarette smoke (Ambrose & Barua, 2004; Brook et al., 2010).

The Atherosclerosis Risk in Communities (ARIC) study looked at null and functional genotypes in *GSTM1* and *GSTT1* in 989 African and European American

participants. Participants who smoked ≥ 20 pack-years and carried the null *GSTT1* genotype had the highest mean levels of CRP, but the smoking interaction term did not reach statistical significance (Miller et al., 2003). The STANISLAS group conducted a similar study using a sample of 265 French individuals.

This group identified smoking interactions between CRP, white blood cell (WBC) count, and TNFα and *GSTM1* genotypes ($p<0.05$); however, they did not observe any significant interactions between smoking and *GSTT1* genotypes for any of the biomarkers measured (Habdous, Siest, Herbeth, Vincent-Viry, & Visvikis, 2004). The relatively limited sample sizes in these studies likely limited power to detect significant interactions.

We sought to assess whether the effect of smoking pack years on CRP is modified by genetic polymorphisms in *GSTM1* and *GSTT1* using a large multi-ethnic consortium of both European-American and African-American participants. Participants came from community-based cohorts from the Candidate Gene and Association Resource (CARe) consortium with genotype data from the ITMAT Broad-CARe (IBC) Chip, a custom 50,000 SNP gene-centric array having dense coverage of over 2,000 candidate genes for CVD pathways.

Results

Characteristics of the 20,874 EA and 6,489 AA study participants can be found in Table 3.1. Smoking prevalence was higher in European Americans, whereas CRP levels and BMI were higher in African Americans. Allele frequencies, by race, for all variants tested can be found in Table 3.2. Large allele-frequency differences were observed for several of the polymorphisms in *GSTT1* and *GSTM1* between EAs and AAs.

Nominal evidence for a main-effect association between log-CRP and *GSTT1* variant rs405597 (p=0.014) was observed in EAs; no other variants in *GSTM1* or *GSTT1* were associated with log-CRP (Table 3.3). A significant interaction between rs405597 and log pack-years (p=0.002) on log CRP was observed in EAs. No other significant interactions were observed in either EAs (Table 3.4) or AAs (Table 3.5). To further understand the interaction between log pack-years and rs405597 genotype on CRP, we performed stratified association analysis, by genotype, ethnicity and cohort (Table 3.6) with log-pack-years as the primary predictor. Among EAs, there was a consistent trend, across cohorts, of a stronger positive relationship between log-pack-years and log-CRP levels for homozygotes carrying the common allele, in relative comparison to the other genotype groups. For the most part, with exception to ARIC, we observed a consistent trend supporting an additive interaction between genotype and log pack-years on log-CRP in EAs. Consistent with the lack of evidence for an interaction in AAs, we saw no evidence for consistent trends regarding the effects of smoking on CRP between cohorts or genotype categories in this population.

To further explore the *GSTT1* gene variants, we performed a haplotype analysis in both our European and African American cohorts. These SNPs define seven haplotypes in EAs and six haplotypes in AAs (freq $\geq$ 0.02). Haplotype analysis in EAs confirmed that rs405597 tags a haplotype associated with elevated CRP levels after adjustment for age, gender, log pack-years, and 10 principal components (p=0.04). Haplotype analysis in AAs for the same haplotype or any other identified haplotypes showed no association (p=0.75).

Discussion

It has been a long-standing hypothesis that mechanisms of detoxification regulate disease susceptibility in cancer research; however, this concept has only recently begun to surface in the field of cardiovascular disease (Conklin & Bhatnagar, 2011). The mechanisms linking detoxification and CVD remain largely unknown except in the case of GSTs (Hayes, Flanagan, & Jowsey, 2005).

Glutathione is a tripeptide composed of glutamine, cysteine, and glycine. The thiol group is a potent reducing agent making it the most abundant intracellular small molecule. Glutathione is an important antioxidant as it plays a primary role in the detoxification of a variety of electrophilic compounds and peroxides via the catalysis of GSTs (Anderson, 1998; Mullineaux et al., 1998). GSTs are multifunctional proteins, meaning in one environment they could protect against disease by removing harmful toxins yet in a different environment they could attenuate the effects of beneficial substances (Conklin & Bhatnagar, 2011). Although there are many classes of GSTs, to our knowledge, the only ones that may contribute to variations in pathology or clinical response are the mu and theta classes (C. M. Smith et al., 1994; Strange et al., 2000).

Several studies have explored the interaction between smoking and the mu and theta classes of GSTs in relation to CVD outcomes. One study assessed the effects of smoking with *GSTT1* and *GSTM1* null genotypes in a case-cohort coronary artery disease (CAD) study. Subjects who were current smokers and carried the *GSTM1* null genotype had a 1.63-fold higher relative risk of CAD, and those smokers who carried the *GSTT1* null genotype had a 2.6-fold higher relative risk (Tamer et al., 2004). Another study

assessed parental effects of cigarette exposure. Cresci, et al. investigated the effect of maternal and paternal cigarette exposure in children who did not produce *GSTM1* and *GSTT1* enzymes in relation to congenital heart disease risk (Cresci et al., 2011). A significant interaction was observed between *GST* genotypes and parental exposure to cigarette smoke; children with null *GST* genotypes had a greater risk than children with functional genotypes when both parents were exposed to cigarette smoke. The ARIC study investigated the role of *GSTT1* and *GSTM1* genotypes modifying the effect of smoking on subclinical CVD events. They observed that smokers with a history of 20 or more pack years who carried *GSTT1* coding-region deletions had increased odds, compared with never smokers and smokers with less than 20 pack years, of having preclinical atherosclerosis, which was measured by carotid IMT (intima-medial thickness) (Olshan et al., 2003).

While previous reports have studied the relationships between polymorphisms in *GSTM1* and *GSTT1* , nicotine exposure and disease risk (Hayes & Strange, 2000)(10-Du 201), no large study has characterized these relationships using common genetic variants (Moyer et al., 2007). The most commonly studied form of *GSTM1* and *GSTT1* variants have been deletion polymorphisms, which have been shown to cause a reduction in *GST* expression and the inability to detoxify xenobiotics and carcinogens (Townsend et al., 2003). The prevalence of null (or deletion) genotypes has been reported to be highest among Chinese and Koreans and lowest among Mexican Americans (Nelson et al., 1995). In the ARIC cohort, included among our cohorts, the prevalence of *GSTT1* deletion polymorphisms is higher among African Americans (~33%) compared to

European Americans and the prevalence of *GSTM1* deletion polymorphisms is higher among European Americans (~50%) compared to African Americans (Miller et al., 2003).

Oxidative stress plays a major role in the pathogenesis of atherosclerosis, a primary risk factor of CVD, and GSTs detoxify metabolites produced by oxidative stress protecting the cell against injury (Doney, Lee, Leese, Morris, & Palmer, 2005; Turkanoglu, Can Demirdogen, Demirkaya, Bek, & Adali, 2010). It has been shown that when GST enzymatic activity is inhibited, oxidative-stress induced apoptosis is exacerbated (Roth et al., 2011). *GSTT1* deletion polymorphisms are linked to cardiovascular diseases because they reduce levels of cellular detoxification due to minimal or no GST expression (Ntais, Polycarpou, & Ioannidis, 2005; Raimondi et al., 2006; White, Li, Nurgalieva, & El-Serag, 2008; Ye, Song, Higgins, Pharoah, & Danesh, 2006). We identified five transcription sites via the TRANSFAC (a transcription factor database) (Wingender et al., 2000) that rs405597 could be altering suggesting its role in *GSTT1* gene expression. Therefore, exposure to nicotine may cause a decrease in GST expression thus reducing glutathione conjugation of electrophilic compounds and peroxides. Consequently, reduction in conjugation results in increased levels of reactive oxygen species (ROS) and the oxidation of lipids in the endothelium (Townsend et al., 2005).

We identified one *GSTT1* polymorphism (rs405597) that showed both nominally significant evidence for a main-effect association and significant evidence for an interaction with smoking on CRP in EAs. Rs405597 is in modest linkage disequilibrium

(LD) with a *GSTT1* deletion (D'=0.29) in CEUs and in strong LD (D'=0.99) in Yorubans (YRI) (Zhao, Marotta, Eichler, Eng, & Tanaka, 2009). It is unclear whether the observed associations with *GSTT1* rs405597 are due to LD with the deletion. Interestingly, we note that we saw no evidence for either main effects or interaction effects with rs405597 in AAs even though this variant is in stronger LD with the deletion in this population.

The major strength of this study is the large sample size of both European and African Americans with measured CRP, smoking exposure data and *GSTT1* and *GSTM1* genotype data. Generalization to other populations remains to be determined and the functional mechanisms linking genetic variants in the genes to CRP concentrations under nicotine exposure are also unknown, as the SNPs studied may not be causally linked to variation in CRP concentration levels but rather be in linkage disequilibrium with the causal variants. We only had a limited number of *GSTM1* and *GSTT1* SNPs for study and it is conceivable that we missed important effects with other variants that could have been detected with broader SNP coverage, including uncommon variants, in and around these genes. Perhaps most importantly, our study was unable to include data on the deletions in these genes that have been previously studied, limiting our ability to draw direct inference into whether our observed interaction between *GSTT1* rs405597 and pack-years reflected linkage disequilibrium with previously studied null variants.

Lastly, we recognize that the use of pack-years as a smoking variable does present potential drawbacks in our analysis. Pack-years is a widely use smoking index in most epidemiological studies because it combines the duration of smoking and smoking magnitude (Hellenbrand et al., 1997; Uchimoto et al., 1999).

However, this index makes the assumption that the dose of smoking one cigarette per day for 10 years is equivalent to smoking 10 cigarettes per day for one year (Doll & Peto, 1978). For the hazard of smoking related diseases such as lung disease, duration of smoking appears to have a greater effect than the number of cigarettes smoked per day (Doll & Peto, 1978). In contrast, in the case of heart disease, smoking intensity appears to have a greater role in disease risk. Disease risk decreases rapidly after initial smoking cessation, but the rate of decreased risk slows down after the first year of cessation (Landmark, 2001). In a recent study that assessed the relationship between four measures of smoking intensity (pack-years, cigarettes/day, the Fagerstrom Test of Nicotine Dependence, and exhaled carbon monoxide levels) and CRP levels, correlations were only observed for pack-years as a smoking index, albeit the correlation was modest ($r=0.05$) (Asthana et al., 2010). The assessment of inflammatory markers such as CRP, it is unclear whether measures of both smoking cessation and duration should be considered (Wannamethee et al., 2005). Finally, strong correlations between CRP and adiposity may mask the effects of smoking on CRP (Asthana et al., 2010). We chose to include BMI as a covariate in our models to partially address any possible confounding between smoking, genotype and smoking on CRP levels. Higher order interaction models including BMI would be of interest, though sample size constraints would likely severely limit power even for our relatively large study.

Our findings illustrate the importance of studying the interaction between genes and environmental exposures on subclinical measures of CVD. Specifically, our results provide supporting evidence regarding the interaction between *GSTT1* genotypes and smoking behavior on CRP. Our results could provide better insight about the pathways

involved in CVD initiation and progression. Moreover, these findings also illustrate the need to create better smoking indices to measure the effects of gene-by-environment interactions. It also alludes to the need to characterize more variants within the *GST* classes so that we can better interrogate these genes, ultimately determining the functionally important variants, to better understand their relationship to cardiovascular disease risk.

Materials and Methods

CARe

The CARe (Candidate Gene Association Resource) consortium consists of nine studies. The purpose of the consortium was to bring together deeply phenotyped prospective cohort studies to increase power for genetic association scans of CVD and other disorders (Musunuru et al., 2012). Cohorts included in this smoking interaction with CRP levels were Atherosclerosis Risk in Communities (ARIC), the Cardiovascular Health Study (CHS), the Framingham Heart Study (FHS), and the Multi-Ethnic Study of Atherosclerosis (MESA). Further details of the participating CARe studies are reported in the **Supplemental Methods.**

Genotyping

CARe

The IBC SNP array is described in detail in Keating et al. (Keating et al., 2008). The IBC SNP array includes 49,320 SNPs selected across ~2,000 candidate loci for CVD.

The array includes SNPs that capture patterns of genetic variation in both European- and African-descent populations. Genotyping for the CARe cohorts was performed at the Broad Institute (Cambridge, MA). Criteria for DNA sample exclusion based on genotype data included sex mismatch, discordance among duplicate samples, or sample call rate <95%.

For each set of duplicates or monozygotic twins, data from the sample with the highest genotyping call rate were retained. SNPs were excluded when monomorphic, the call rate was <95%, or HWE was $p<10^{-5}$ in EAs. Given the genetic admixture in African Americans, there was no HWE filter used for these samples.

Data analysis

Participants with CRP measurements over 10 mg/L were excluded from analysis as these observations would be highly influential and potentially a result of acute infection. We natural log-transformed CRP level to generate an approximately normal distribution of model residuals, conditional on the covariates, to meet linear model assumptions. We assumed an additive genetic model in all tests of association and modeled log-transformed pack-years to assess the impact of smoking. Main-effect (between log-CRP and genotype, with smoking as a covariate) and interaction-effect (modeling interaction between genotype and smoking on log-CRP) association analyses were performed using linear regression models implemented in PLINK (Purcell et al., 2007a) for cohorts with unrelated individuals and linear mixed models implemented in GWAF for cohorts with related individuals (M. H. Chen & Q. Yang, 2010). All models were stratified by cohort and race.

We adjusted for age, sex, body mass index (BMI), LDL-C, HDL-C, and the first 10 principal components calculated using the program EIGENSTRAT (A. L. Price et al., 2006) to control for potential population substructure. After obtaining cohort-specific results, we performed fixed-effects, inverse variance-weighted meta-analyses, separately by race, using the METAL software (C. J. Willer, Y. Li, & G. R. Abecasis, 2010).
A statistical threshold of $\alpha = 0.0045$ (0.05/11) was thus used to declare significance.
Finally, we performed selected haplotype-based association analyses around interesting variants using the haplo.stats package for R (Lake et al., 2003).

Table 3.1. Participant characteristics by cohort/ethnicity

	ARIC		CHS		FHS	JHS	MESA	
	EA	**AA**	**EA**	**AA**	**EA**	**AA**	**EA**	**AA**
N	8574	2600	3766	692	6472	1864	2035	1333
Female, %	53.9	64.1	56.3	62.5	53.8	60.6	52.0	53.7
Pack-years	15.04±22.94	8.59±19.52	18.63±27.3	13.76±23.8	7.73±16.83	29.6±28.6	14.37±27.4	11.42±18.7
Age, y	54.1±5.7	52.9±5.7	72.8±5.6	73.0±5.7	48.8±13.7	50.0±12.0	62.2±10.1	61.7±9.8
Body Mass Index, kg/m2	26.9±4.7	29.6±5.9	26.4±4.5	28.5±5.6	27.4±5.5	32.2±7.8	27.7±5.1	30.1±5.7
CRP, mg/L	4.02±5.6	5.8±7.1	4.3±6.7	6.1±8.2	3.3±6.2	5.1±7.7	3.4±5.1	4.7±6.9
LDL, mg/dL	137.8±37.8	138.3±43.3	129.9±35.6	129.2±36.6	139.7±38.0	126.4±37.1	117.1±30.1	116.8±32.6
HDL, mg/dL	50.5±16.7	54.8±17.4	53.5±15.7	57.9±15.5	54.1±16.5	51.6±14.8	52.5±15.8	52.4±15.1

Table 3.2. Allelic frequencies of *GSTM1* and *GSTT1* polymorphisms in European and African Americans

				European Americans			***African Americans***		
Chromosome	**SNP**	**Gene**	**Location**	**Major**	**Minor**	**MAF**	**Major**	**Minor**	**MAF**
22	rs405597	GSTT1	22734578	A	C	0.06	A	C	0.11
22	rs422674	GSTT1	22736778	C	A	0.36	C	A	0.07
22	rs738809	GSTT1	22735492	A	G	0.28	A	G	0.30
22	rs11090305	GSTT1	22737483	T	C	0.19	T	C	0.40
1	rs574344	GSTM1	110015037	T	A	0.08	T	A	0.10
1	rs655315	GSTM1	110016701	A	G	0.49	A	G	0.19
1	rs12024479	GSTM1	110021609	C	G	0.48	C	G	0.28
1	rs7537275	GSTM1	110022262	T	A	0.001	T	A	0.13
1	rs412543	GSTM1	110031467	C	G	0.04	C	G	0.14
1	rs7553593	GSTM1	110048621	G	A	0.0004	G	A	0.02
1	rs10857797	GSTM1	10050887	T	G	0.02	T	G	0.37

Table 3.3. Main effects for polymorphisms in *GSTM1* and *GSTT1* in European Americans and African Americans for C-reactive protein

		European Americans			***African Americans***		
SNP	**Gene**	**Beta**	**Se**	**P-value**	**Beta**	**Se**	**P-value**
rs405597	GSTT1	-0.052	0.021	0.014	-0.031	0.031	0.319
rs422674	GSTT1	0.005	0.011	0.643	0.004	0.038	0.918
rs738809	GSTT1	0.009	0.011	0.434	-0.008	0.021	0.692
rs11090305	GSTT1	-0.007	0.013	0.616	-0.008	0.020	0673
rs574344	GSTM1	-0.026	0.018	0.163	0.082	0.033	0.012
rs655315	GSTM1	-0.005	0.010	0.653	0.026	0.025	0.290
rs12024479	GSTM1	-0.005	0.010	0.628	0.014	0.022	0.514
rs7537275	GSTM1	0.042	0.161	0.796	-0.020	0.029	0.488
rs412543	GSTM1	0.061	0.025	0.015	-0.059	0.060	0.327
rs7553593	GSTM1	-0.04	0.232	0.919	-0.005	0.061	0.929
rs10857797	GSTM1	-0.0007	0.035	0.984	-0.002	0.026	0.951

Table 3.4. Interaction effects with log-pack years for polymorphisms in *GSTM1* and *GSTT1* in European Americans

SNP	**Gene**	**Main effect Beta**	**Main effect SE**	**Interaction Beta**	**Interaction SE**	**Interaction Pvalue**
rs405597	GSTT1	-0.107	0.027	0.041	0.013	0.002
rs422674	GSTT1	0.004	0.014	0.001	0.007	0.889
rs738809	GSTT1	0.004	0.015	0.002	0.007	0.802
rs11090305	GSTT1	-0.0101	0.017	0.002	0.008	0.802
rs574344	GSTM1	-0.032	0.024	0.005	0.011	0.639
rs655315	GSTM1	-0.0006	0.013	-0.0003	0.006	0.960
rs12024479	GSTM1	-0.016	0.013	0.006	0.006	0.339
rs7537275	GSTM1	0.060	0.202	-0.042	0.099	0.674
rs412543	GSTM1	0.069	0.033	-0.005	0.015	0.761
rs7553593	GSTM1	-0.165	0.273	-0.098	0.201	0.625
rs10857797	GSTM1	0.005	0.047	-0.006	0.021	0.765

Table 3.5. Interaction effects with log-pack years for polymorphisms in *GSTM1* and *GSTT1* in African Americans

SNP	**Gene**	**Main Effect Beta**	**Main Effect SE**	**Interaction Beta**	**Interaction SE**	**Interaction Pvalue**
rs405597	GSTT1	-0.024	0.039	-0.006	0.021	0.777
rs422674	GSTT1	-0.002	0.048	0.005	0.024	0.838
rs738809	GSTT1	-0.009	0.026	0.002	0.014	0.873
rs11090305	GSTT1	-0.002	0.025	-0.009	0.013	0.500
rs5760176	GSTT1	0.028	0.034	-0.016	0.017	0.366
rs574344	GSTM1	0.056	0.041	0.021	0.022	0.327
rs655315	GSTM1	-0.01	0.031	0.031	0.016	0.049
rs12024479	GSTM1	0.010	0.028	0.001	0.014	0.920
rs7537275	GSTM1	-0.059	0.037	0.032	0.019	0.094
rs412543	GSTM1	-0.103	0.077	0.035	0.039	0.370
rs7553593	GSTM1	0.082	0.078	-0.077	0.042	0.067
rs10857797	GSTM1	0.021	0.032	-0.019	0.017	0.278

Table 3.6. Main effects for association with log-pack years and C-reactive protein by genotype for rs405597

	rs405597_AA		rs405597_AC		rs405597_CC	
European Americans	Beta(SE)	P-value	Beta(SE)	P-value	Beta(SE)	P-value
ARIC	0.092(0.008)	$2.1x10^{-32}$	0.028(0.021)	0.194	0.038(0.148)	0.015
CHS	0.072(0.009)	$1.5x10^{-14}$	0.032(0.028)	0.254	-0.040(0.218)	0.858
FHS	0.144(0.010)	$4.8x10^{-35}$	0.119(0.031)	$1.5x10^{-4}$	n/a	n/a
MESA	0.050(0.014)	$2.1x10^{-4}$	-0.051(0.041)	0.217	-0.119(0.244)	0.646
African Americans						
ARIC	0.107(0.020)	$6.3x10^{-8}$	0.061(0.040)	0.132	0.388(0.196)	0.069
CHS	0.082(0.026)	0.002	0.071(0.062)	0.253	0.329(0.315)	0.406
JHS	0.078(0.020)	$9.9x10^{-5}$	0.081(0.044)	0.067	0.003(0.207)	0.989
MESA	0.061(0.020)	0.002	0.135(0.040)	$8.8x10^{-4}$	0.380(0.198)	0.151

CHAPTER 4

GENOME-WIDE ASSOCIATION STUDY OF AFRICAN AMERICANS AND HOMOCYSTEINE LEVELS

Cardiovascular disease (CVD) is the leading cause of death globally; 1 in 4 deaths in the US are caused by CVD. African Americans (AAs) have the highest rates of CVD and CVD mortality in the US; approximately 46% of AAs have been diagnosed with CVD (Henry et al., 2012). Plasma homocysteine (Hcy) is an established marker for CVD and elevated Hcy concentrations are associated with venous thrombosis, coronary heart disease, stroke, and atherosclerosis (Cattaneo, 1999; Homocysteine Studies, 2002; Wald et al., 2002; Welch & Loscalzo, 1998). Because it is influenced by multiple environmental factors (Brattstrom et al., 1994), its role in CVD is interesting and complex. Age, gender, renal function, vitamin intake, cigarette smoking, and menopause are among the main environmental determinants of Hcy levels (Brattstrom et al., 1994). Hcy also has an important genetic component, as heritability is estimated to be 47%-70% (Lange et al., 2010) and genetic mutations in enzymes for Hcy metabolism have long been known to be associated with Hcy level. There are at least two functional variants (677C>T and 1298A>C) in the gene for methylenetetrahydrofolate reductase (*MTHFR*), which catalyzes the conversion of 5,10-methylenetetrahydrofolate to 5-methyltetrahydrofolate, the primary circulating form of folate (Brustolin, Giugliani, &

Felix, 2010). Mutations in the gene for cystathione-β-synthase (*CBS*), which catalyzes the transsulfuration converting Hcy to cystathionine, which in turn is converted to cysteine, is known to markedly increase Hcy level and leads to increased rates of premature stroke and venous thrombosis (Selhub, 1999).

Genome-wide association (GWA) studies have been conducted for Hcy, but these have so far been limited to populations of European ancestry (Hazra et al., 2009; Malarstig et al., 2009; Pare et al., 2009; Tanaka et al., 2009) and Filipinos (Lange et al., 2010). All of these GWA studies observed evidence for association with Hcy for common variants in the *MTHFR* and *CBS* loci. A GWA study in 13,974 EA women from the Women's Genome Health Study (WGHS) identified and replicated two novel loci for Hcy: *DPEP1* and *CPS1,* although *CPS1* was replicated only in women in the confirmatory sample. The *CPS1* locus was also found to be associated with Hcy in a GWA study of 1786 Filipino women and additional analysis in the adult offspring of the original sample confirmed that the association is only in women. Evidence for modification of genotype effects by sex has also been seen for *MTHFR*, where the effect magnitude of the 677C>T polymorphism is stronger in males than females (Lange et al., 2010; Pare et al., 2009).

While no GWA studies for Hcy in samples of African ancestry have been reported to date, candidate gene studies have been conducted in AA samples. The Coronary Artery Risk Development in Young Adults (CARDIA) performed a candidate gene study to assess the associations between variants in *MTHFR*, *CBS* and two

additional genes for Hcy enzymes before and after folic acid fortification, which was mandated by the US Food and Drug Administration in 1996. Significant associations were detected in both European and AAs pre-folic acid fortification era at all loci; however, post-folic era, no associations at the *MTHFR* variants were detected in AAs (Tsai et al., 2009).

A study about thhe clinical utility of the *MTHFR* 677 C>T variant reported that the prevalence of homozygote (TT) genotypes were much smaller in AAs versus EAs (<1% compared to 12% respectively) (Tsai et al., 2009). In an analysis of *MTHFR* polymorphisms conducted in the WGHS, no associations between either of the known *MTHFR* SNPs (677C>T and 455G>A) and Hcy were observed in AA women (Albert et al., 2009). Additionally, the prevalence of *MTHFR* genotypes was lowest in AA women (~4 %) in a study that assessed the heterogeneity of *MTHFR* polymorphisms in women from four different ethnic backgrounds (Esfahani et al., 2003).

Herein, we performed a genome-wide association meta-analysis of AA samples from three cohorts to identify loci associated with Hcy concentration. Study participants came from CARDIA, the Jackson Heart Study (JHS), and the Multi-Ethnic Study of Atherosclerosis (MESA).

Results

Characteristics of the 3,817 African American study participants are presented in Table 1. Consistent with prior reports, Hcy levels are higher in men. A higher proportion of men currently smoke while women have, on average, higher BMI.

The Manhattan plot for men is shown in Figure 1a and the Quantile-Quantile (QQ) plot is shown in Figure 4.2a. In men, we observed genome-wide significant evidence for an association with Hcy at chromosome 8 SNP rs7834157 ($p=4.9x10^{-8}$), which maps 30kb upstream of *C8orf4* (Figure 4.3b). No evidence supporting this association was observed in women (p= 0.54) (Table 4.2). In women (Figure 4.1b), our top result, chromosome 14 SNP rs11628917 ($p= 9.8x10^{-8}$), which maps ~1kb downstream of *SERPINA1,* narrowly missed genome-wide significance. No evidence for association was found for this same SNP in men (p= 0.54).

In our combined-sex meta-analysis we identified significant evidence for association at chromosome 8 SNP rs341697 ($p = 1.9x10^{-8}$), which maps ~80kb downstream of *CSMD1* in a region not previously reported to be associated with Hcy levels. Additionally, we identified two chromosome 11 polymorphisms (rs2289123, $p=4.1x10^{-9}$; rs317194 $p=1.5x10^{-8}$) significantly associated with Hcy levels in the *NOX4* gene (Figure 1c). SNP rs2289123 maps <100bp upstream of the five prime untranslated region of *NOX4* while rs317194 is an intronic SNP.*4*. Polymorphisms in *NOX4* were previously reported to be associated with Hcy levels in a genome-wide analysis of EA women conducted by the Women's Genome Health Study (Pare et al., 2009).

Conditional Analysis

We performed a conditional analysis in the regions surrounding rs7834157, identified in our male cohorts, and rs341697, identified in our combined sex datasets (See Figures 4.3a and 4.3b). For both regions, after conditioning on the top SNP in the region, the results for the remaining SNPs were attenuated (p> 0.005).

Comparison with other previously identified GWA loci

To compare our results to loci previously identified to be associated with Hcy in other GWA studies, we used log transformed Hcy values to maintain consistency with the models used in previous studies (Table 4.3).

The well-studied non-synonymous polymorphism, rs1801133, in *MTHFR* at chromosome 1 only reached nominal significance (p=0.028) in our dataset. Evidence of sex specific effects was observed for non-synonymous polymorphism rs2274976 in the same locus region. The effect was significant in women (p=0.038), but not significant and in the opposite direction in men (p=0.77). *CBS*, which is involved in the transsulfuration pathway, was nominally associated (rs6586282, p=0.02) with Hcy in our sex-combined dataset. Previously reported variants in *NOX4* and *MUT* had the most significant associations (rs11018628 at *NOX4*, p= 1.1×10^{-4}; rs4267943 at *MUT*, p= 9.6×10^{-4}). We also observed evidence supporting the association between variant rs7422339 at the *CPS1* gene and Hcy in women (p=0.033). Consistent with earlier reports (Lange et al., 2010; Pare et al., 2009), no evidence for such an association was observed in men (p=1.00).

Association of CSMD1 rs341697 and IDO2 rs7834157 with other CVD-related traits

We tested the association between variants rs341697 at *CSMD1* and rs7834157 at *C8orf4(TC1)* against LDL-C, HDL-C, BMI, and log C-reactive protein. For both variants, we observed no associations with any of the CVD related traits even before or after adjustment for Hcy levels (data not shown).

Discussion

To our knowledge, this study conducted in approximately 3000 AAs from three community-based cohorts represents the first GWA scan for Hcy concentration in individuals of African descent. We identified novel associations at 8p23 (rs341697) near *CSMD1* in the sex-combined meta-analysis, and at 8p11 (rs7834157) near *C8orf4(TC1)* in the male-only meta-analysis. We also observed significant evidence for association at *NOX4*, a gene previously identified in a GWA analysis in EA women from the Women's Genome Health Study (WGHS).

The cub and sushi multiple domains 1 (*CSMD1*) gene is a regulator of complement activity (Kraus et al., 2006) and has been implicated in several diseases, including hypertension and Kawasaki disease (Burgner et al., 2009; Hong et al., 2010). The Korean Association Resource (KARE) study performed a GWA analysis to identify variants associated with hypertension and blood pressure using approximately 8800 subjects of Korean descent, followed by a replication analysis of the top SNPs in participants from the Health2 project (a Korean cohort of approximately 7800 individuals). They identified a significant association at SNP rs995322 ($p<0.001$) near *CSMD1* for hypertension (Hong et al., 2010). A GWA analysis using an initial sample of Dutch individuals and a second sample of Australian trios to identify loci associated with Kawasaki Disease, an inflammatory vasculitis, which predominantly affects young children, observed association at SNP rs2912272 ($p<0.002$) at *CSMD1 (Burgner et al., 2009)*.

We observed significant evidence for association at rs7834157 near the gene

TC1(C8orf4), specific to males. *TC1* is well conserved among vertebrates suggesting an important biological role(Y. Kim et al., 2006). *TC1* has been implicated to have a potential role in inflammatory and immune regulation (J. Kim et al., 2009). *TC1* is up-regulated by the cytokine IL-1β, TNF-α, and cellular stresses in various cells (Y. Kim et al., 2006; Park et al., 2007). *TC1* also enhances the proliferation of dendritic cells, which are essential for B cell development (Y. Kim et al., 2006; Lee et al., 2005). A study by Kim, et al. demonstrated that *TC1* up-regulated the expression of IL-6, IL-1α, and ICAM1 in human aortic endothelial cells (HAEC). Other adhesion molecules were also up-regulated (COX-2, CCL5, CXCL1, VCAM1, and E-selectin). The transfection of *TC1* in human umbilical vein endothelial cells (HUVECs) also up-regulated IL-6, IL-1α, and ICAM1 expression (J. Kim et al., 2009). Knock down of *TC1* by shTC1 in HAECs caused the expression of IL-6, IL-1α, and ICAM1 to be significantly down-regulated indicating the requirement of *TC1* in endothelial inflammatory gene regulation. In addition, the effect of I-kB-kinase (IKK), an NF-kB inhibitor, suppressed *TC1* expression as well as expression of IL-1β regulated *TC1* expression suggesting that *TC1* expression is dependent on NF-kB activity. NF-kB is activated under inflammatory conditions promoting the release of cytokines, chemokines, and intercellular adhesion molecules in endothelial cells (J. Kim et al., 2009). *TC1* also appears to have clinical relevance with respect to vascular regulation. *TC1* downstream genes have been implicated in atherosclerosis (Hansson & Libby, 2006; Pober & Sessa, 2007; Rader & Daugherty, 2008), and there is evidence of *TC1* down-regulation by cardiovascular protective agents such as resveratrol and epigallocatechin-3-gallate (EGCG). Resveratrol, found in red wine, and EGCG, a catechin in green tea, have been shown to have significant anti-

inflammatory activities (Opie & Lecour, 2007; Potenza et al., 2007; Shenouda & Vita, 2007). Overall, previous literature suggests that this gene may have implications in CVD.

The most associated SNP in the current study (rs2289123) is at *NOX4*, an enzyme in the nicotinamide adenine dinucleotide phosphate (NADPH) oxidase (Nox) family, and was originally identified to be associated with Hcy in the WGHS. The Noxes are a family of enzymes that serve as a significant source of cellular reactive oxygen species (ROS). Low to moderate levels of ROS have been shown to contribute to important functions such as apoptosis, cell differentiation, adhesion, and senescence; however it has been demonstrated that elevated levels of ROS are associated with a variety of diseases including CVD (Lambeth, Kawahara, & Diebold, 2007). The *NOX4* gene is expressed in a wide array of tissue and cell types, which include cardiovascular tissues, hematopoetic stem cells, adipocytes, epithelial cells, and podocytes (Ikeda et al., 2011). Unlike the other four Noxes that make up the Nox family, *NOX4* does not require the binding of cytosolic proteins for ROS production but instead produces ROS constitutively (Martyn, Frederick, von Loehneysen, Dinauer, & Knaus, 2006). *NOX4* expression in vascular cells was documented after its initial discovery in the kidneys (Geiszt, Kopp, Varnai, & Leto, 2000). Studies support that *NOX4* is the most abundant Nox isoform in the vasculature. In blood vessels, *NOX4* has a greater than 1000-fold copy number increase over *NOX1* and *NOX2* and its expression has also been reported to be dominant in human and mouse cardiac myocytes (Byrne et al., 2003). Similar to ROS, elevated expression of *NOX4* has been seen in patients with atherosclerosis, hypertension, and ischemic stroke (F. Chen, Haigh, Barman, & Fulton, 2012). The amount of ROS contributed by NOX4 is regulated

at the transcriptional level (F. Chen et al., 2012) and the most significant SNP in this region (rs2289123) in our study is in moderate linkage disequilibrium (D'=0.7) with an *NOX4* variant (rs2289124) in the 5' untranslated region. Data on this variant (rs2289124) comes from 1000 Genomes and we imputed SNPs from HapMap Phase 2 data, thus rs2289124 was not covered in our analysis. However, this observation suggests that the combined polygenic effects of these variants may have a role in regulating *NOX4* expression in vascular cells and tissues.

In human atherosclerosis, *NOX4* expression is increased in intimal lesions of coronary arteries (Sorescu et al., 2002). Decreased *NOX4* expression has been seen in laminar flow, which has been shown to be protective against atherosclerosis and increased expression in oscillatory shear stress, which is proatherogenic. Furthermore, a variety of oxidized lipids, such as ox-LDL can stimulate *NOX4* expression in macrophages, which can then be scavenged by receptors to form atherosclerotic lesions. Because *NOX4* can be influenced by multiple factors such as the severity of the lesion, the location of the blood vessel, and the presence of diseases such as inflammation and heart failure (Guzik et al., 2006), its role in atherogenesis still remains indeterminate. Similar to *NOX4*, the thiol metabolite of Hcy can combined with LDL-cholesterol to produce aggregates that are susceptible to uptake by macrophage (Kang, 1995); however the exact mechanisms underlying this gene's influence on Hcy levels remains unclear.

Interestingly, both rs341697 near *CSMD1* and rs7834157 near *TC1* lie within or near regions of structural variation specifically regions consisting of copy number variants (CNVs). Data from previously reported CNVs tend to overstate the genomic

extent of actual variation while underestimating variation among individuals. Therefore, these data could implicate that genomic alterations of these genes are frequent and do not necessarily predispose to disease risk (Shaikh et al., 2009); however, analysis of copy number variations across the human genome by Shaikh, et al. found these findings to be inconsistent. Studies from the Database of Genetic Variation (DGV) collectively report 49 CNVs (mean size: 347kb) within *CSMD1* including seven segmental duplications spanning large stretches of the gene and 5 additional CNVs predicted to disrupt one or more *CSMD1* exons (Conrad, Andrews, Carter, Hurles, & Pritchard, 2006; Perry et al., 2008; Sharp et al., 2005). In contrast, Shaikh, et al. identified 507 CNVs (mean size: 7535bp) and only four were predicted to disrupt exonic sequences. Moreover, this group did not identify any of the segmental duplications previously reported suggesting the possibility that these CNVs could be rare variants (Shaikh et al., 2009). The CSMD1 SNP observed in the current study (rs341697), lies within one of the CNV regions reported in the Shiakh study. Four loss of function CNVs identified in a study by the 1000 Genomes Project consortium were downstream of *C8orf4(TC1)* rs7834157 (Genomes Project et al., 2010). Thus, these variants could potentially be tagging the effects of nearby CNVs on Hcy concentration.

Although variants in the *MTHFR* gene were the first to be implicated in the regulation of Hcy levels (Gudnason et al., 1998), their estimated effects in AAs in this study were notably smaller than seen in populations of European and Asian ancestry (Heijmans et al., 1999; Lange et al., 2010; Pare et al., 2009) which is consistent with what has been observed in other studies of AAs (Albert et al., 2009; Johnson et al., 2010).

While the association at *CPS1* did not reach genome-wide significance, we observed nominally significant evidence of female-specific effects at rs7422339. The association of *CPS1* with Hcy was initially identified in women from the WGHS were the evidence for replication in that study came only from women in the follow-up sample, and the Cebu Longitudinal Health and Nutrition Survey (CLHNS) GWA in Filipinos also observed evidence for this locus only in women. The consistency of these findings suggest that variants in or near *CPS1* may play a role in the regulation of homocysteine in women but have little or no effect in men.

We also observed nominal evidence of female-specific effects at *MTHFR* , which is in contrast to evidence observed for male-specific effects observed in the CLHNS and several studies of populations of European ancestry (Pare et al., 2009; Tanaka et al., 2009). It is possible that the smaller effects of *MTHFR* may only be observed because *MTHFR* genotypes appear to be dependent on folate status, and generally African American women have lower levels with respect to EA men and women and AA men (Perry, et al).

Strengths of this study include that this represents the first GWA study for Hcy in AAs, and the relatively large sample size and high-density of common variants through genotype imputation allowed us to identify new genes influencing Hcy not previously identified in GWA studies of EAs and Filipinos. A limitation of this study is that we have not obtained confirmation of the two novel loci in an independent AA sample. These results, however, do provide new candidates for other AA samples with measured Hcy level. As with any GWA study, SNPs identified may not be causally linked to variation

in Hcy concentrations, rather they may be in linkage disequilibrium with the causal variants. Even in the case that some of the variants identified may be functional, the functional mechanisms linking them to Hcy concentrations remain to be determined.

Our findings provide additional insight regarding the genes that may be involved in the regulation of Hcy levels across and within populations. While the *NOX4* locus appears to have similar effects between AAs and EAs on Hcy concentration, the effects of *MTHFR* loci appear to be considerably smaller in AAs. In addition, given that the two novel loci identified in this study were not reported in the WGHS GWA of almost 14000 EAs suggest that there may be genes with AA-specific effects on Hcy. Our findings emphasize the need to conduct GWA studies in non-European populations.

Materials and Methods

Study samples and Hcy measurement

CARDIA

CARDIA is a longitudinal study of the evolution of coronary heart disease risk, started in 1985-86 in 5,115 AA and EA men and women, then aged 18-30 years. The CARDIA sample was recruited was recruited at random during 1985-86 primarily from geographically based populations in Birmingham AL, Chicago IL, and Minneapolis MN and, in Oakland, CA, from the membership of the Kaiser-Permanente Health Plan. Examinations after baseline were year 2 (1987-88, n=4624, 90% retention, year 5 (1990-91, n=4352, 85% retention), year 7 (1992-93, n=4086, 80% retention), year 10 (1995-96, n=3950, 79% retention), year 15 (2000-2001, n=3672, 74% retention) and year 20 (2005-06, n=3549, 72% retention).

Fasting serum samples that were collected in 2000 and stored frozen at −70 °C were used for Hcy measurement. tHcy was measured in serum by a fluorescence polarization immunoassay (IMx Homocysteine Assay, Axis Biochemicals ASA, Oslo, Norway) using the IMx Analyzer (Abbott Diagnostics, Abbott Park, IL).

JHS

The Jackson Heart Study (JHS) is a prospective population-based study to seek the causes of the high prevalence of common complex diseases among African Americans in the Jackson, Mississippi metropolitan area. During the baseline examination period (2000-2004) 5,301 self-identified African Americans were recruited from four sources, including (1) randomly sampled households from a commercial listing; (2) ARIC participants; (3) a structured volunteer sample that was designed to mirror the eligible population; and (4) a nested family cohort. Unrelated participants were between 35 and 84 years old, and members of the family cohort were ≥ 21 years old when consent for genetic testing was obtained and blood was drawn for DNA extraction.

Fasting blood samples were collected and assayed for Hcy. Hcy was measured at the University of Minnesota, in conjunction with the University of Minnesota Medical Center, Fairview. Plasma Hcy was assayed by the fluorescence polarization immunoassay method, supplied by Abbott Diagnostics (Abbott Park, IL).

MESA

MESA is a cohort study of the characteristics of subclinical cardiovascular disease and the risk factors that predict progression to clinically overt cardiovascular disease or progression of subclinical disease. MESA researchers study a diverse, population-based

sample of 6,814 asymptomatic men and women aged 45-84: 2,622 white (39%), 1,893 African-American (28%), 1,496 Hispanic (22%) and 803 (12%) of Asian (primarily Chinese) descent. Participants were recruited from six Field Centers across the United States (Winston-Salem, NC; St. Paul, MN; Chicago, IL; Los Angeles, CA; New York, NY; Baltimore, MD).

Blood samples collected at baseline and were processed with the use of a standardized protocol and stored at 80 °C until analyzed. Participants were asked to fast for 12 hours, avoid smoking on the morning of the exam, and avoid heavy Exercise 12 hours before the exam. Total plasma Hcy was measured with high-performance liquid chromatography with fluorometric detection. The coefficient for variation for this assay was 3.8%.

Genotyping/ Genotyping/Imputation

Genome-wide genotyping on all cohorts was performed using the Affymetrix 6.0 SNP Array (Affymetrix, Santa Clara, Calif). Genotype data was originally available on 906,600 SNPs across the genome. SNPs were excluded if they had success rate less than 90% of samples; subjects were removed if they had success rate less than 95% of SNPs. No SNPs were removed due to significant deviation from Hardy–Weinberg equilibrium (HWE) expectations because the AA population is an admixed population, which may result in departures from HWE expectations even under ideal conditions. After quality control, genotype data was available on ~800,000 autosomal SNPs in 3,817 AA subjects with Hcy measures.

Genotype imputation was performed using the software MACH (Y. Li et al., 2010). HapMap phase 2 data sets (www.hapmap.org) were used for the reference panel (build 36 release 22) consisting of both YRI and CEU HapMap-phased haplotypes. Genotype imputation resulted in allelic-dosage data, representing the expected number of copies of the minor allele a subject carries, on approximately 2.2 million autosomal SNPs.

Data analysis

Hcy levels were highly skewed and simple transformations (e.g. natural log) did not remedy numerous outliers. Hence, we used inversed-normalized residuals, obtained from modeling Hcy on established non-genetic risk factors: age, body mass index (BMI) and current smoking status, as our outcome measure to meet linear model assumptions of normally distributed residuals and to minimize influence of extreme values when modeling Hcy on genotype. Residuals were calculated separately by cohort and by sex. We systematically performed linear regression models, implemented in PLINK (Purcell et al., 2007b),to test for associations between Hcy (using normalized residuals) and individual SNPs, with covariate adjustment for the first 10 principal components calculated using the program EIGENSTRAT (A.L. Price et al., 2006) to control for potential population substructure, assuming an additive genetic model. The association analyses were performed separately by cohort and by sex. After obtaining specific cohort results, we performed fixed-effect, inverse variance-weighted meta-analyses across cohorts using the METAL software (C.J. Willer et al., 2010) first separately by sex and then on the combined samples of men and women. Based on convention for imputed

GWA studies, a p-value threshold of $\square$=5.0x10^{-8} was used as the significance threshold for our results to maintain an overall experimental type I error rate of ~5%. Finally, conditional analyses, where association analyses are repeated with additional covariate adjustment for the top SNP in the region, were performed in selected regions that demonstrated evidence for association to assess whether there was any evidence for multiple independent signals in the regions of interest.

Table 4.1. Descriptive table of participants by study

	CARDIA		JHS		MESA	
	Men	***Women***	***Men***	***Women***	***Men***	***Women***
N	167	284	796	1220	623	727
Current Smoker, %	31.4	27.8	19.5	12.2	20.0	15.8
Age, y	24.70±3.75	24.60±3.87	49.49±11.84	50.26,12.05	61.78±9.88	61.57±9.82
BMI, kg/m^2	24.63±3.77	26.08±5.80	30.2±6.42	33.49±8.25	28.79±4.51	31.28±6.28
Homocysteine, umol/L	10.38±3.08	8.33±2.61	9.77±3.03	8.67±4.23	10.38±4.91	8.92±3.21

Table 4.2. Evidence of most significant associations with Hcy values for stratified and combined meta-analysis

						Both		*Men*		*Women*	
Chr	**GENE**	**SNP**	**Location**	**Effect Allele**	**MAF**	**Beta(SE)**	**Pvalue**	**Beta(SE)**	**Pvalue**	**Beta(SE)**	**Pvalue**
11	*NOX4*	rs2289123	88864366	T	0.29	0.16(0.028)	4.09×10^{-9}	0.20(0.043)	3.47×10^{-6}	0.14(0.037)	1.59×10^{-4}
11	*NOX4*	rs317194	88727291	T	0.43	-0.16(0.029)	1.45×10^{-8}	-0.16(0.045)	1.78×10^{-4}	-0.16(0.037)	2.09×10^{-5}
8	*CSMD1*	rs341697	2729496	T	0.31	-0.15(0.028)	1.90×10^{-8}	-0.14(0.044)	9.58×10^{-4}	-0.16(0.036)	5.10×10^{-6}
8	*C8orf4 (TC1)*	rs7834157	40093249	C	0.17	0.12(0.032)	7.95×10^{-5}	0.27(0.050)	4.88×10^{-8}	0.03(0.042)	0.539
14	*SERPINA1*	rs11628917	93913472	T	0.05	-0.29(0.063)	3.05×10^{-6}	-0.06(0.108)	0.544	-0.42(0.078)	9.77×10^{-8}

Table 4.3. Evidence of association with log-Hcy values at previously reported GWA loci

				Original Study		*Current Study*						
							Both		*Men*		*Women*	
Author	**Chr**	**GENE**	**SNP**	**MAF**	**Beta**	**MAF**	**Beta(SE)**	**Pvalue**	**Beta(SE)**	**Pvalue**	**Beta(SE)**	**Pvalue**
Lange, LA, et al. (W)	1	*MTHFR*	rs1801133	0.22	0.083	0.12	0.028(0.01)	0.028	0.023(0.02)	0.239	0.032(0.02)	0.060
Tanaka, et al. (M and W)	1	*MTHFR*	rs1801133	0.47	1.31	0.12						
Pare, G, et al. (W)	1	*MTHFR*	rs2274976	0.05	-0.062	0.05	-0.026(0.02)	0.137	0.009(0.03)	0.768	-0.046(0.02)	0.038
Lange, LA, et al. (W)	2	*CPS1*	rs7422339	0.24	0.076	0.32	0.013(0.008)	0.100	-0.000(0.01)	0.998	0.022(0.01)	0.033
Pare, G, et al. (W)	2	*CPS1*	rs7422339	0.31	0.027	0.32						
Pare, G, et al. (W)	11	*NOX4*	rs11018628	0.07	-0.05	0.14	0.036(0.009)	1.1×10^{-4}	0.034(0.01)	0.021	0.038(0.01)	0.002
Pare, G, et al. (W)	6	*MUT*	rs4267943	0.36	0.024	0.39	0.022(0.007)	9.6×10^{-4}	0.014(0.01)	0.169	0.028(0.009)	0.001
Pare, G, et al. (W)	16	*DPEP1*	rs459920	0.44	0.024	0.22	0.012(0.009)	0.214	0.002(0.01)	0.811	0.022(0.01)	0.100
Pare, G, et al. (W)	21	*CBS*	rs6586282	0.18	-0.030	0.25	-0.021(0.009)	0.023	-0.02(0.01)	0.071	-0.018(0.01)	0.160

**for comparison purposes these betas are based on log Hcy*

Figure 4.1a. GWAS results for inverse normal Hcy levels in 1, 586 African American men. *Loci described in the text are annotated with gene symbols*

.

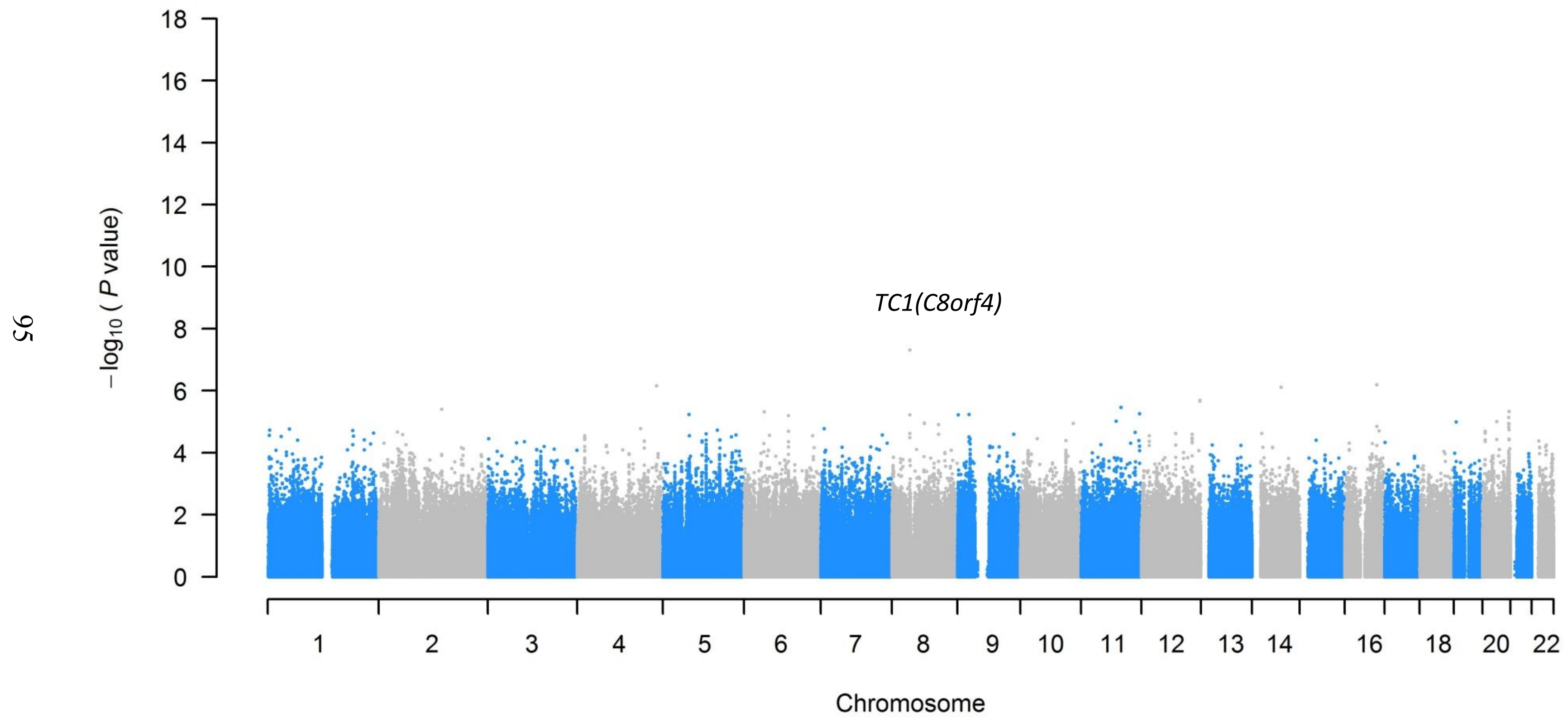


Figure 4.1b. GWAS results for inverse normal Hcy levels in 2,231 African American women. *Loci described in the text are annotated with gene symbols*

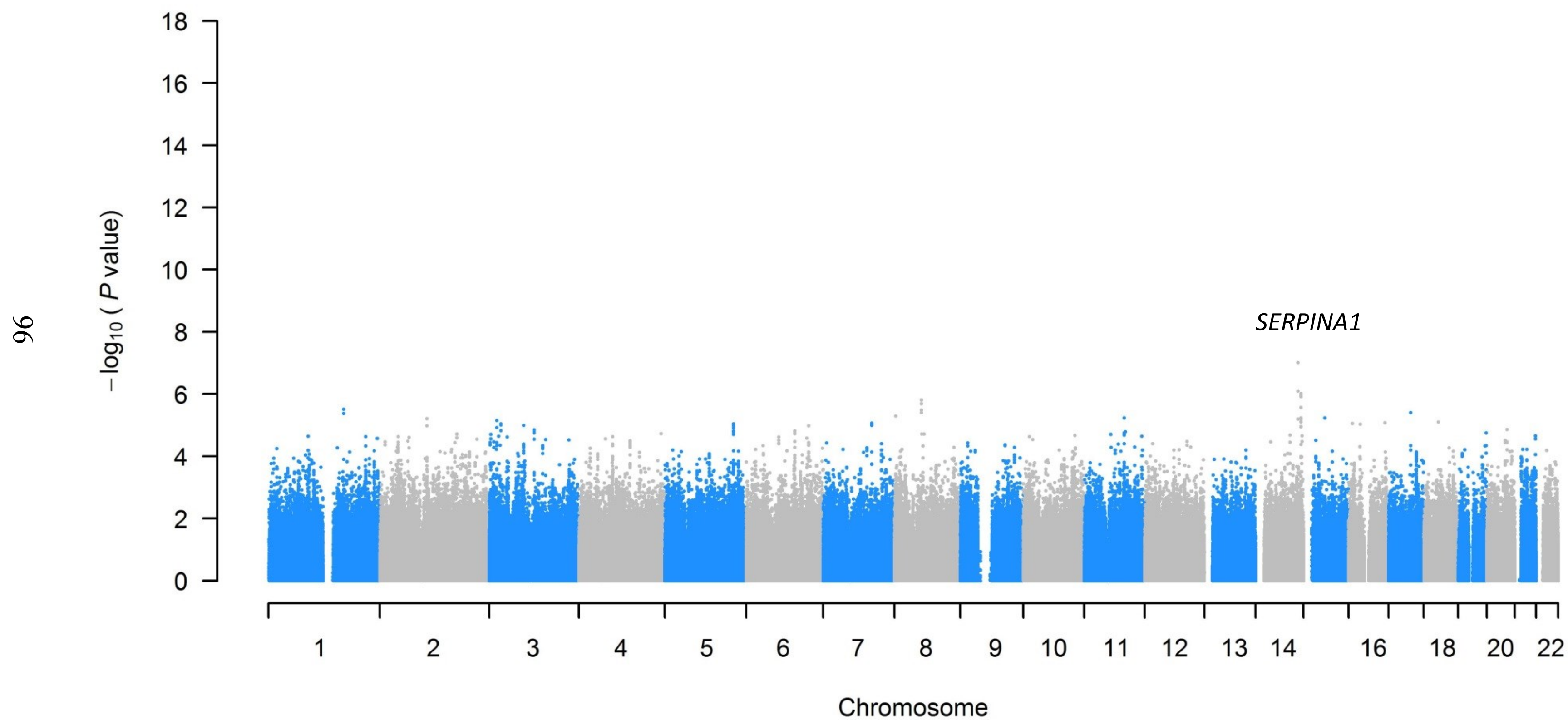


Figure 4.1c. GWAS results for inverse normal Hcy levels in 3,817 African American men and women. *Loci described in the text are annotated with gene symbols.*

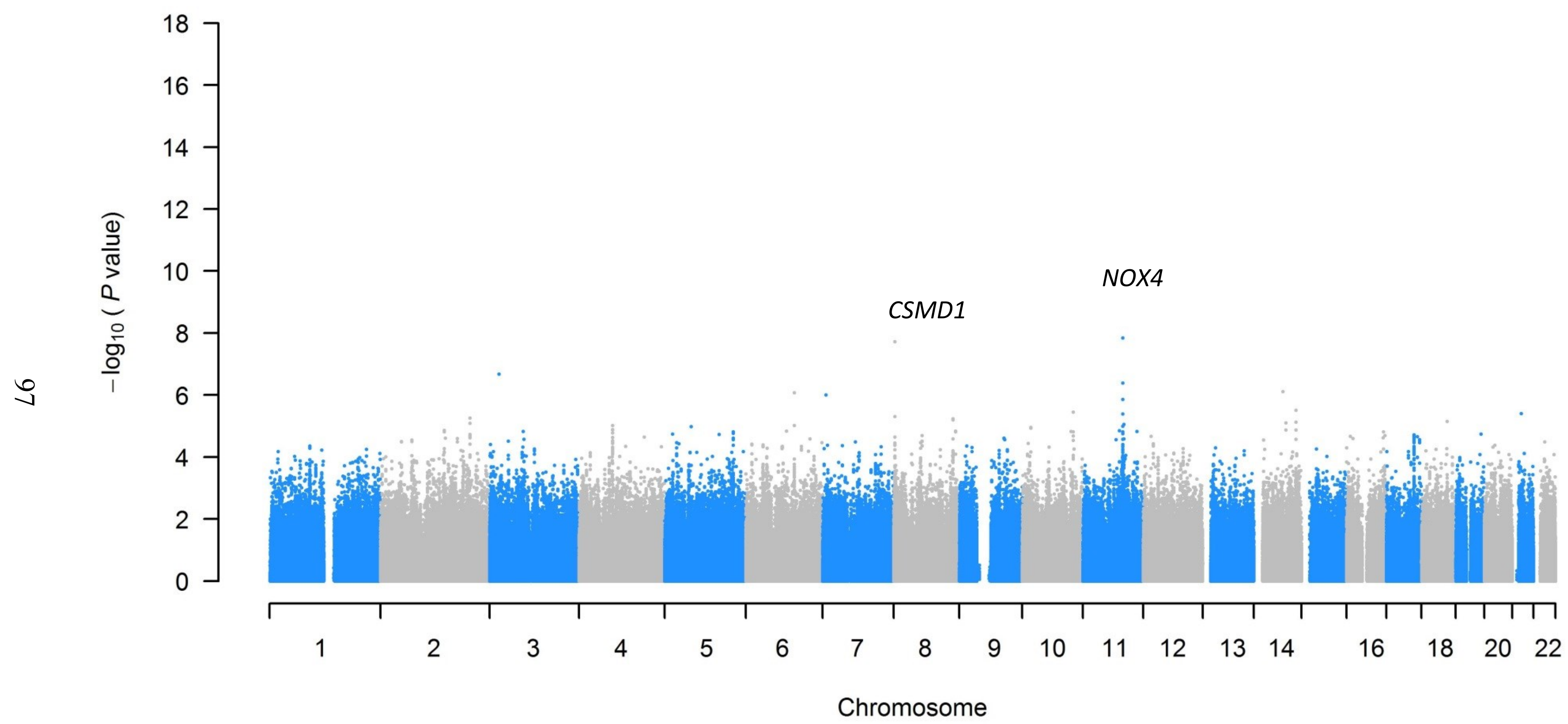


Figure 4.2a. Quantile-quantile plot of SNP association with inverse-normalized Hcy values.in African American males.

Red circles indicate all SNPs. A solid black line indicates the uniform expectations and the gray dashed lines indicate the 95% prediction interval

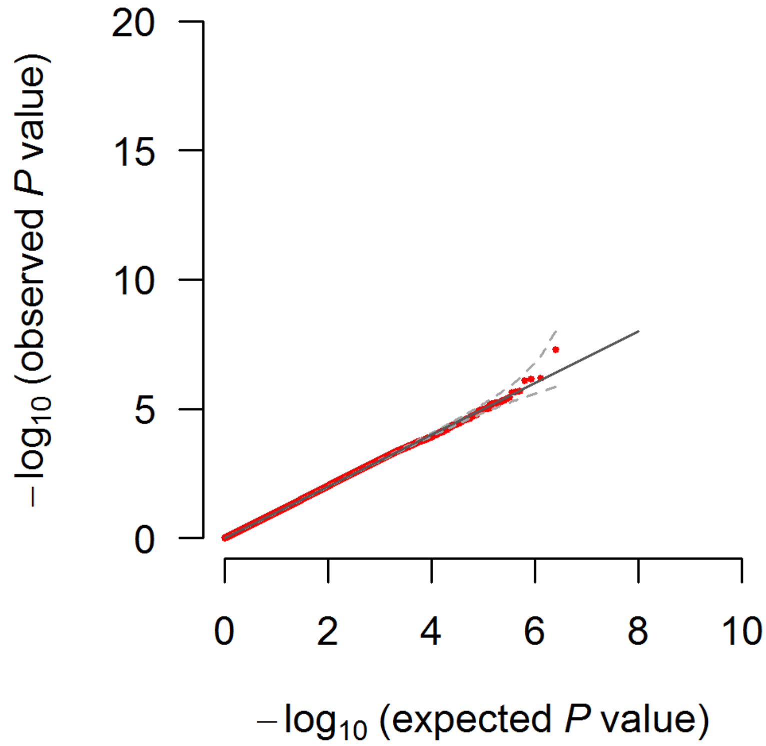


Figure 4.2b. Quantile-quantile plot of SNP association with inverse-normalized Hcy values.in African American females.

Red circles indicate all SNPs. A solid black line indicates the uniform expectations and the gray dashed lines indicate the 95% prediction interval

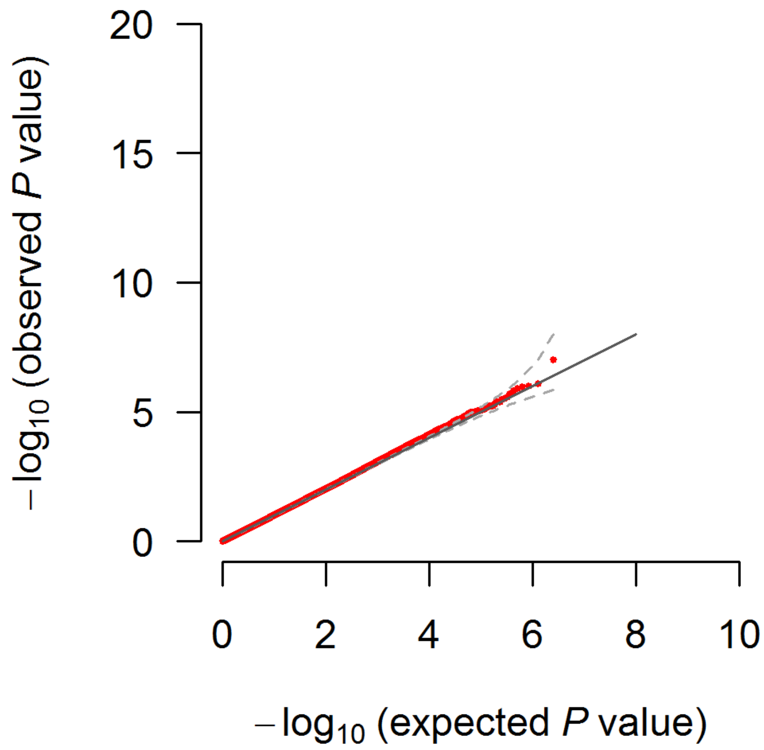


Figure 4.2c. Quantile-quantile plot of SNP association with inverse-normalized Hcy values.in African American males and females.

Red circles indicate all SNPs. A solid black line indicates the uniform expectations and the gray dashed lines indicate the 95% prediction interval

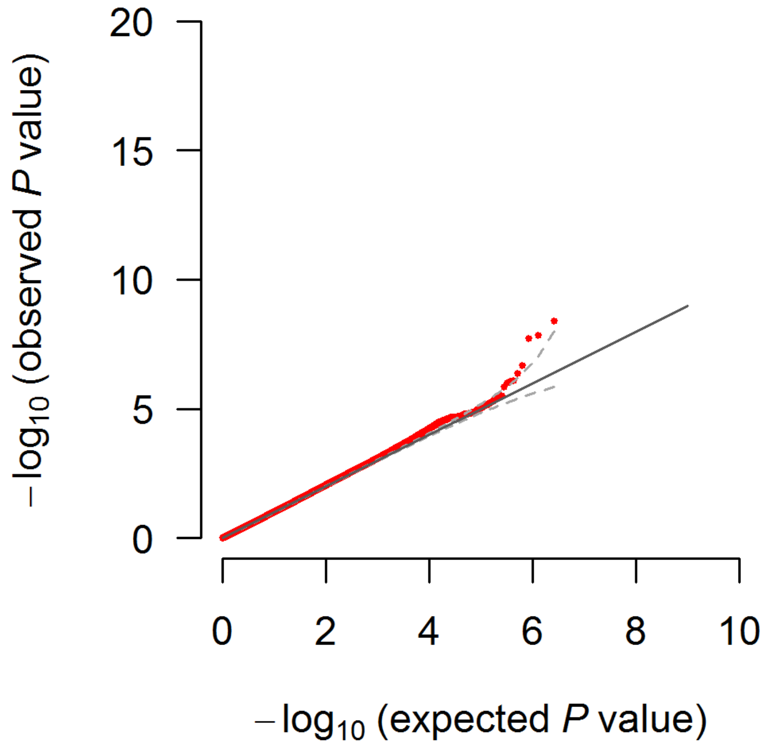


Figure 4.3a. Locus zoom plot of the *CSMD1* gene region (8p11) in African American men and women.

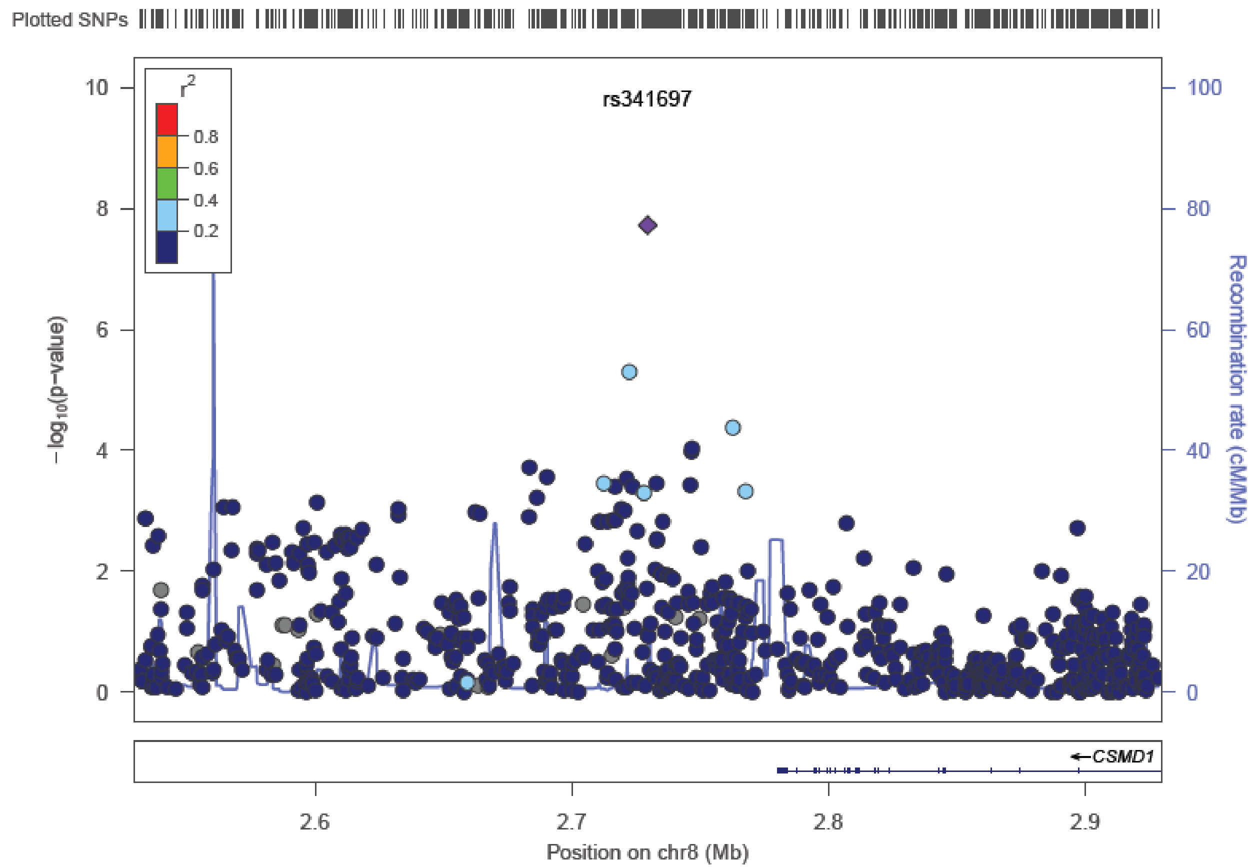
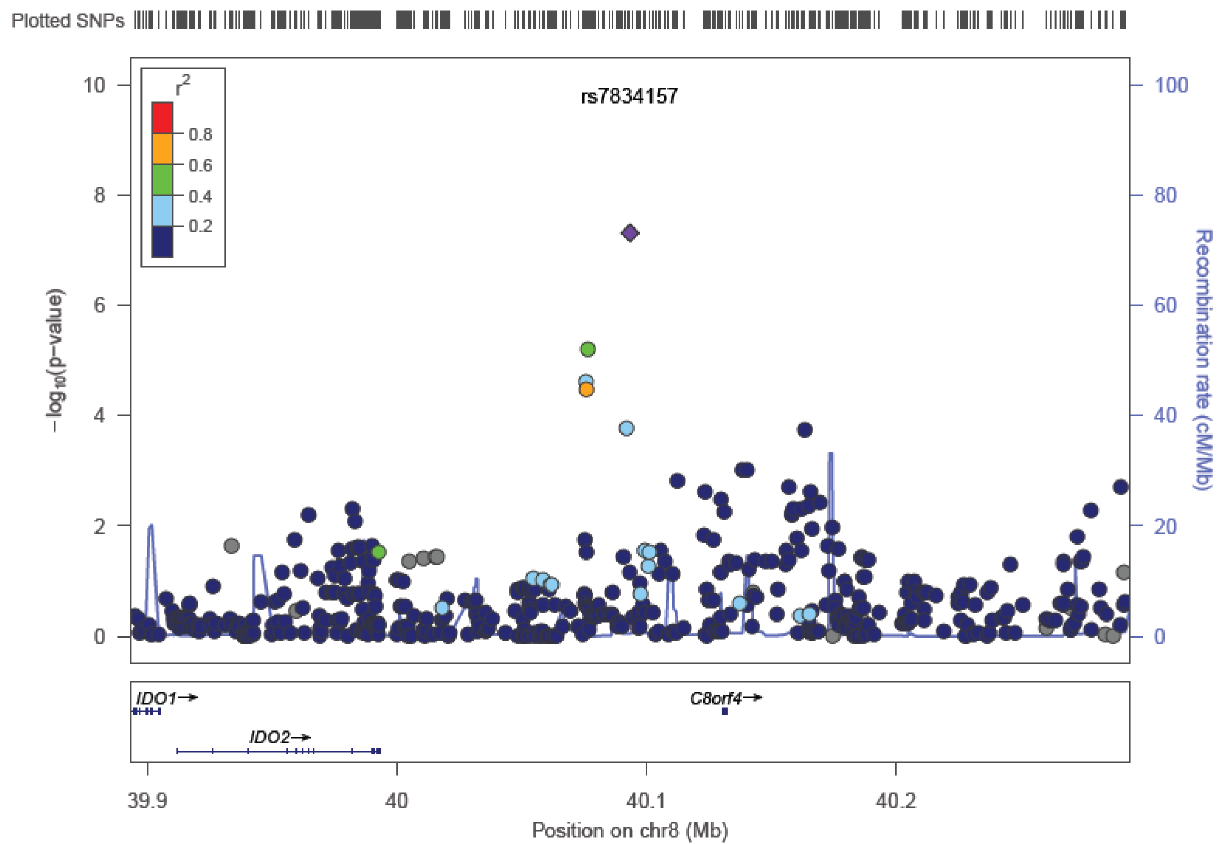


Figure 4.3b. Locus zoom plot of the *TC1(C8orf4)* gene region (8p23) in African American men



CHAPTER 5

CONCLUSIONS

The aims of this dissertation were to identify loci that influence CRP and Hcy levels and assess whether two *GST* gene variants modify the effects of cigarette smoking on CRP. The first paper identified common and uncommon variants associated with CRP in EAs and AAs using a dense SNP array designed specifically to capture the variation of candidate genes for CVD pathways (IBC chip). The second paper focused on testing for interactions between polymorphisms in two genes involved in cellular detoxification, *GSTM1* and *GSTT1*, and cigarette smoking on CRP level in EAs and AAs. The third paper was a genome-wide association scan conducted on AAs from three population-based cohorts to identify loci associated with Hcy levels.

Major Findings

The results of the first paper support a novel association between an uncommon *CD36* variant and CRP in AAs. It also provides additional confirmatory evidence for several previously identified genes associated with CRP in both EAs and AAs. The non-synonymous SNP rs3211938 identified to be associated with CRP is specific to populations of African ancestry, and the uncommon variant was associated with lower CRP levels. This is in line with previous literature that supports the association between this SNP and elevated HDL levels conferring a cardio-protective effect in AAs.

The results from the second paper provide support that *GSTT1* genotype modifies the association between cigarette smoking and CRP in EAs, with log-pack-years of smoking being more positively associated with CRP level in individuals who

are homozygous for the more common allele as compared to the other two genotypes. Thus, the data support that the uncommon variant for SNP rs405597 is protective of the effects of amount of cigarette smoking, as assessed by log-pack-years. While this analysis utilized polymorphisms available on the IBC chip, it is likely that a functional deletion, tagged by the rs405597 SNP is responsible for the effect modification.

Lastly, we identified two putative novel associations in a genome-wide association scan for Hcy: at 8p23 near *CSDM1* and a male-specific locus at 8p11 near *TC1(C8orf4)*. We also replicated evidence for the *NOX4* locus, which was identified in EA women from the Women's Genome Health Study, and sex-specific effects at *CPS1* in AA women, which was previously observed in EAs and Filipinos. Interestingly, there was little evidence of effect for two polymorphisms in the well-studied *MTHFR* gene established as influencing Hcy in populations of European and Asian ancestry. The results from this study underscore the value of conducting genome-wide association studies in non-European populations.

While genome-wide association and candidate gene studies have been successful in identifying multiple loci associated with CVD, most of these studies have been focused on populations of European descent. It has been demonstrated that the use of non-European samples can improve the power to detect susceptibility loci (N.A. Rosenberg et al., 2010). Pulit et al. showed that minor allele frequency differences across populations

can lead to dramatically different levels of statistical power to detect associations (Pulit et al., 2010). Because of these differences in allele frequencies, the incorporation of multiple ethnic samples can help elucidate the range of genetic loci influencing phenotypes of interest. Our studies provide strong empirical evidence supporting the value of using other populations to identify loci associated with CVD traits. Notably, we identified an uncommon variant, near private to African Americans, in *CD36* that is associated with CRP, a gene that has not been previously implicated to be important to CRP. We also identified two novel loci for Hcy in the first GWA study of Hcy in AAs, including one near the strong candidate gene *CSDM1*.

Future Directions/Studies

Chapters 1 and 3 described large-scale association studies used to identify loci associated with CRP and Hcy, respectively, while paper 2 described a targeted candidate gene study designed to assess whether the effect of exposure to cigarette smoke on CRP is modified by variants in genes previously reported to interact with smoking exposure.

For chapter 1, future replications studies using similar populations would be desirable to corroborate our observed finding of an association between CRP and a *CD36* variant rs3211938. While we have confirmed the finding in our own replication sample, future studies could provide refined and more precise effect-size estimates. This variant is specific to populations of African descent; previous literature has identified an association between rs3211938 and HDL levels in African Americans, suggesting that this variant has a protective effect in this population (Love-Gregory et al., 2011). While

evidence for association in our own study remained after covariate adjustment for HDL levels, we only had one HDL measurement available to us that was measured at the same time or prior to the CRP measurement for most samples. Larger sample sizes that include repeated measures of CRP and HDL would be valuable in further understanding the relationship between this variant, HDL and CRP. Our findings certainly suggest that this and other *CD36* variants warrant further investigation regarding their predictive value for clinical endpoints and utility as a diagnostic tool to help clinicians determine CVD risk in AAs.

Chapter 2 describes a candidate gene study used to assess whether *GSTM1* and *GSTT1* genotypes modify the effects of smoking exposure on CRP. Larger samples and more precise smoking indices will be necessary to confirm our identified interaction between cigarette smoking and *GSTT1* rs405597 and CRP. We used pack-years as a measure of smoking index in our analysis. This measure has been widely used in many epidemiological studies because it combines measures of smoking magnitude and intensity. However, we understand that this index lacks precision because it makes the assumption that the dose of smoking one cigarette per day for 10 years is equivalent to smoking 10 cigarettes per day for one year (Doll & Peto, 1978). Second, size of the pack can also differ from brand to brand but there is a general agreement that there are 20 cigarettes in one pack. Ideally, multiple factors should be taken into account when measuring smoking status, such as levels of carbon monoxide exhalation, types of cigarettes smoked (filter and non-filter), age and start of passive smoking (environmental tobacco exposure) (Indrayan & Sarmukaddam, 2001). Because accurately quantifying all

these measures presents itself with tremendous difficulty, it is more important to identify the most optimal measure for the outcome being investigated.

Additionally, this study also alludes to the need to characterize more variants within the *GST* classes so that we can more fully capture the variation in these genes, including functionality important less common variants. Much of the literature is focused on *GST* deletion regions and it would be useful to more precisely understand the relationship between our common variants and these deletions. Further analysis of LD patterns and haplotypes in these genes across multiple populations may also uncover novel functionally critical variants.

For chapter 3, replication studies are necessary to confirm evidence of association between Hcy and genes *CSMD1* and *TC1(C8orf4)* identified in our genome-wide analysis of AAs and whether these findings can be generalized to other populations. Further, the sex-specific result of *TC1* being only seen in AA males needs to be validated, and if true, better understood. It would be a valuable exercise to fine map these regions surrounding these loci using haplotype analyses and next-generation sequencing data to identify any uncommon variants that our leading SNPs may be tagging. In addition, because we observed that both rs7834157 (*TC1(C8orf4)*) and rs341697 (*CSMD1*) are near regions of copy number variation (CNVs), we will utilize available copy number variation data to determine whether these SNPs are tagging the effects of these known CNVs Both genes have a biological role in relation to CVD. *CSMD1* is located in a region that contains a large amount of immune-related genes and is hypervariable in humans (Burgner et al., 2009). *CSDM1* has been implicated in hypertension in a Korean cohort (Hong et al.,

2010) and Kawasaki disease, an inflammatory vasculitis, which predominantly affects young children, in a Dutch cohort (Burgner et al., 2009).

TC1(C8orf4) has been implicated in diverse biological regulations. There is evidence of *TC1* down-regulation by cardiovascular protective agents such as resveratrol and epigallocatechin-3-gallate (EGCG) suggestive of its clinical relevance in vascular regulation (Opie & Lecour, 2007; Potenza et al., 2007; Shenouda & Vita, 2007). Pharmacological inhibition of NF-κB caused a downregulation of *TC1* expression in human aortic endothelial cells (HAECs) supporting evidence that *TC1* expression is dependent on NF-κB activity (J. Kim et al., 2009). However, further investigation is needed to determine the roles of *TC1* and *CSMD1* in the regulation of Hcy levels.

Public Health Implications

CVD is the leading cause of death in the US, making it a critical public health concern. Currently, 1 in 3 Americans have some form of heart disease, including high blood pressure, coronary heart disease, heart failure, stroke and other conditions. By 2030, approximately 116 million people in the US (40.5 percent) will have some form of CVD (Heidenreich et al., 2011). In 2010, the total cost of CVDs in the US was estimated to be $444 billion, and productivity losses due to heart disease, which include days missed from work tasks due to illness or potential lost earnings due to premature death, were estimated to be $172 billion (Kuklina, Tong, George, & Bansil, 2012).

New projections from the American Heart Association propose that by 2030, the cost of medical care for CVD-related morbidities will escalate to $818 billion and productively

losses will reach $276 billion (Heidenreich et al., 2011). Although numerous advances have been made in risk assessment, disease prognosis, and the development of more effective therapeutics, there are still critical areas of CVD etiology that require additional insight to continue making progress in reducing the disease burden, including identifying factors that may underlie differences in risk across ethnic groups. Multi-ethnic genetics studies can, in part, address this area of need.

A better understanding of the genetic factors influencing CVD biomarkers across ethnicities, as well as identifying genes that modify the effects of established risk factors on these biomarkers, will provide important insight regarding the pathophysiology of CVD, provide health-care providers with better methods of identifying individuals at increased risk, as well as allowing for the potential of future personalization of interventions and treatment.

APPENDIX I

Supplemental Methods:

Participating cohorts

Atherosclerosis Risk in Communities Study (ARIC)

ARIC is a longitudinal cohort study of atherosclerosis and its clinical sequelae. From 1987 to 1989, a population-based sample of 15,792 men and women aged 45 to 64 years were recruited from 4 US communities (Forsyth County NC, Jackson MS, suburban Minneapolis MN, and Washington County MD). CRP was assessed using the immunoturbidimetric CRP-Latex (II) high-sensitivity assay from Denka Seiken (Tokyo, Japan). This assay, which has been validated against the Dade Behring method (Deerfield, Ill), was performed according to the manufacturer's protocol and using a BN2 analyzer (Dade Behring (Deerfield, Il). To assess repeatability of measurements, 421 blinded replicates were measured on different dates. The reliability coefficient was 0.99

Coronary Artery Risk Development in Young Adults (CARDIA)

CARDIA is a longitudinal study of the evolution of coronary heart disease risk, started in 1985-86 in 5,115 AA and EA men and women, then aged 18-30 years. The CARDIA sample was recruited was recruited at random during 1985-86 primarily from geographically based populations in Birmingham AL, Chicago IL, and Minneapolis MN and, in Oakland, CA, from the membership of the Kaiser-Permanente Health Plan. Examinations after baseline were year 2 (1987-88, n=4624, 90% retention, year 5 (1990-91, n=4352, 85% retention), year 7 (1992-93, n=4086, 80% retention), year 10 (1995-96,

n=3950, 79% retention), year 15 (2000-2001, n=3672, 74% retention) and year 20 (2005-06, n=3549, 72% retention).

At the Year 15 and 20 examinations, bloods for measurement of CRP were collected in 2-mL blue top (citrated) vacutainer tubes and centrifuged at 4°C for approximately 20 minutes. Citrated plasma was promptly separated from cells, transferred to a cryovial, and frozen at −70°C. Frozen samples were shipped to the University of Vermont where CRP was measured using high-sensitivity nephelometry-based methods (BNII nephelometer, Dade Behring). The assay range was 0.175–1100 mg/L, (intrassay CVs, 2.3–4.4%; inter-assay CVs, 2.1–5.7%). Measurement of CRP was blind to depressive symptom scores.

Cleveland Family Study (CFS)

The Cleveland Family Study (CFS) comprises 2,534 individuals (46% AA) from 352 families examined every 4 years over a 16 year period (1990-2006). The study was begun with the initial aim to quantify the familial aggregation of sleep apnea. Index probands (n=275) were recruited from 3 area sleep centers if they had a confirmed diagnosis of sleep apnea and at least 2 first-degree relatives available to be studied. In the first 5 study years, neighborhood control probands (n=87) with ≥ 2 living relatives available were also recruited. All available first-degree relatives and spouses of the case and control probands were recruited. Blood was sampled and DNA isolated for subjects seen in the last 2 exam cycles (n=1447). The 4th exam included 736 subjects (60% African American), with oversampling subjects who had had a microsatellite genome scan. CRP was measured during the 4th exam.

Cardiovascular Health Study (CHS)

The CHS is a population-based, observational study of risk factors for clinical and subclinical cardiovascular diseases. The study recruited participants 65 years and older from 4 US communities (Forsyth County, North Carolina; Sacramento County, California; Washington County, Maryland; and Pittsburgh, Pennsylvania) in 2 phases: 5201 participants from 1989 to 1990, and 687 (primarily AA participants) from 1992 to 1993. Blood was drawn in all participants and samples were promptly centrifuged at 3000g for 10 minutes at 4°C. Aliquots of plasma were stored in a central laboratory at -70°C. CRP was measured in all stored baseline plasma samples by a high-sensitivity immunoassay, with an inter-assay coefficient of variation of 6.25%.

Framingham Heart Study (FHS)

The FHS started in 1948 with 5209 randomly ascertained participants from Framingham, MA, who had undergone biannual examinations to investigate cardiovascular disease and its risk factors. In 1971, the offspring cohort (comprised of 5124 children of the original cohort and the children's spouses) and in 2002, the third generation (consisting of 4095 children of the offspring cohort), were recruited. FHS participants in this study are of European ancestry. CRP was measured in fasting serum samples using a high-sensitivity assay (Dade Behring BN100). The minimum detectable does of this assay is 0.16mg/L, with a standard curve range of 0.16-1000mg/L. The intra-assay coefficient of variation was 3.2%, while the inter-assay coefficient of variation was 5.3%. The final population for this analysis included individuals (Offspring n=3,582, Third Generation n=3,047).

Jackson Heart Study (JHS)

The Jackson Heart Study (JHS) is a prospective population-based study to seek the causes of the high prevalence of common complex diseases among African Americans in the Jackson, Mississippi metropolitan area. During the baseline examination period (2000-2004) 5,301 self-identified African Americans were recruited from four sources, including (1) randomly sampled households from a commercial listing; (2) ARIC participants; (3) a structured volunteer sample that was designed to mirror the eligible population; and (4) a nested family cohort. Unrelated participants were between 35 and 84 years old, and members of the family cohort were ≥ 21 years old when consent for genetic testing was obtained and blood was drawn for DNA extraction. CRP was measured using the immunoturbidimetric CRP-Latex assay from Kamiya Biomedical Company following manufacturer's high-sensitivity protocol. The inter-assay coefficients of variation on control samples repeated in each assay were 4.5% and 4.4% at CRP concentrations of 0.45 mg/L and 1.56 mg/L respectively. The reliability coefficient for masked quality control replicates was 0.95 for the CRP assay.

Multi-ethnic Study of Atherosclerosis

MESA is a cohort study of the characteristics of subclinical cardiovascular disease and the risk factors that predict progression to clinically overt cardiovascular disease or progression of subclinical disease. MESA researchers study a diverse, population-based sample of 6,814 asymptomatic men and women aged 45-84: 2,622 white (39%), 1,893 African-American (28%), 1,496 Hispanic (22%) and 803 (12%) of Asian (primarily

Chinese) descent. Participants were recruited from six Field Centers across the United States (Winston-Salem, NC; St. Paul, MN; Chicago, IL; Los Angeles, CA; New York, NY; Baltimore, MD). High sensitivity CRP was determined by BNII nephelometer (N High Sensitivity CRP, Dade Behring Inc, Deerfield, IL). The lower limit of detection was 0.17 mg/L.

Women's Health Initiative (WHI)

WHI is one of the largest (n=161,808) studies of women's health ever undertaken in the U.S. There are two major components of WHI: (1) a Clinical Trial (CT) that enrolled and randomized 68,132 women ages 50 – 79 into at least one of three placebo-control clinical trials (hormone therapy, dietary modification, and calcium/vitamin D); and (2) an Observational Study (OS) that enrolled 93,676 women of the same age range into a parallel prospective cohort study. A diverse population including 26,045 (17%) women from minority groups were recruited from 1993-1998 at 40 clinical centers across the U.S. The design has been published. For the CT and OS participants enrolled in WHI and who had consented to genetic research, DNA was extracted by the Specimen Processing Laboratory at the Fred Hutchinson Cancer Research Center (FHCRC) using specimens that were collected at the time of enrollment in to the study (between 1993 and 1998).

According to a standardized protocol, fasting blood specimens were collected from each participant at baseline and processed locally into separate aliquots containing serum, plasma, and buffy coat. The aliquots were frozen and then shipped to a central repository, where they were kept for long-term storage at –70°C. All biochemical assays were carried out by laboratory staff blinded to case/control status (in Dr Rifai's

laboratory). Blood samples from cases and their matched controls were handled identically, shipped in the same batch, and assayed in random order in the same analytical run to reduce systematic bias and interassay variation. High-sensitivity C-reactive protein (hsCRP) was measured on a chemistry analyzer (Hitachi 911; Roche Diagnostics, Indianapolis, Indiana) using an immunoturbidimetric assay with reagents and calibrators (Denka Seiken Co Ltd, Niigata, Japan). The coefficients of variation were 1.61% for hsCRP.

KORA

KORA stands for Cooperative Health Research in the Augsburg Region ("**Ko**operative Gesundheitsforschung in der **R**egion **A**ugsburg"). CRP was measured in EDTA plasma by a high sensitivity in-house immunoradiometric assay (IRMA) in MONICA/KORA S3, using a five-point calibration with WHO International Reference Standard 85/506. The assay range was 0.05-10mg/L. Sample with concentrations >10mg/L were remeasured at higher dilutions. CRP concentrations were determined in triplicate, and the mean was used for analysis. The inter-assay CV for CRP over all ranges was 12%.

REFERNCES

Adams, M. R., Jessup, W., & Celermajer, D. S. (1997). Cigarette smoking is associated with increased human monocyte adhesion to endothelial cells: reversibility with oral L-arginine but not vitamin C. *J Am Coll Cardiol, 29*(3), 491-497.

Ageno, W., Becattini, C., Brighton, T., Selby, R., & Kamphuisen, P. W. (2008). Cardiovascular risk factors and venous thromboembolism: a meta-analysis. *Circulation, 117*(1), 93-102. doi: 10.1161/CIRCULATIONAHA.107.709204

Aitman, T. J., Cooper, L. D., Norsworthy, P. J., Wahid, F. N., Gray, J. K., Curtis, B. R., . . . Scott, J. (2000). Malaria susceptibility and CD36 mutation. *Nature, 405*(6790), 1015-1016. doi: 10.1038/35016636 [doi]

Albert, M. A., Pare, G., Morris, A., Rose, L., Buring, J., Ridker, P. M., & Zee, R. Y. (2009). Candidate genetic variants in the fibrinogen, methylenetetrahydrofolate reductase, and intercellular adhesion molecule-1 genes and plasma levels of fibrinogen, homocysteine, and intercellular adhesion molecule-1 among various race/ethnic groups: data from the Women's Genome Health Study. *Am Heart J, 157*(4), 777-783 e771. doi: 10.1016/j.ahj.2008.12.012

Ambrose, J. A., & Barua, R. S. (2004). The pathophysiology of cigarette smoking and cardiovascular disease: an update. *J Am Coll Cardiol, 43*(10), 1731-1737. doi: 10.1016/j.jacc.2003.12.047

Anderson, M. E. (1998). Glutathione: an overview of biosynthesis and modulation. *Chem Biol Interact, 111-112*, 1-14.

Ansar, W., & Ghosh, S. (2013). C-reactive protein and the biology of disease. *Immunol Res*. doi: 10.1007/s12026-013-8384-0

Assoian, R. K., & Marcantonio, E. E. (1996). The extracellular matrix as a cell cycle control element in atherosclerosis and restenosis. *J Clin Invest, 98*(11), 2436-2439. doi: 10.1172/JCI119059

Asthana, A., Johnson, H. M., Piper, M. E., Fiore, M. C., Baker, T. B., & Stein, J. H. (2010). Effects of smoking intensity and cessation on inflammatory markers in a large cohort of active smokers. *Am Heart J, 160*(3), 458-463. doi: 10.1016/j.ahj.2010.06.006

Auer, P. L., Johnsen, J. M., Johnson, A. D., Logsdon, B. A., Lange, L. A., Nalls, M. A., . . . Li, Y. (2012). Imputation of exome sequence variants into population- based samples and blood-cell-trait-associated loci in African Americans: NHLBI GO

Exome Sequencing Project. *Am J Hum Genet, 91*(5), 794-808. doi: 10.1016/j.ajhg.2012.08.031

Auer, P. L., Srivastava, S., & Doerge, R. W. (2012). Differential expression--the next generation and beyond. *Brief.Funct.Genomics, 11*(1), 57-62. doi: elr041 [pii];10.1093/bfgp/elr041 [doi]

Austin, M. A., Edwards, K. L., McNeely, M. J., Chandler, W. L., Leonetti, D. L., Talmud, P. J., . . . Fujimoto, W. Y. (2004). Heritability of multivariate factors of the metabolic syndrome in nondiabetic Japanese americans. *Diabetes, 53*(4), 1166-1169.

Badimon, L., & Vilahur, G. (2012). LDL-cholesterol versus HDL-cholesterol in the atherosclerotic plaque: inflammatory resolution versus thrombotic chaos. *Ann N Y Acad Sci, 1254*, 18-32. doi: 10.1111/j.1749-6632.2012.06480.x

Bahekar, A. A., Singh, S., Saha, S., Molnar, J., & Arora, R. (2007). The prevalence and incidence of coronary heart disease is significantly increased in periodontitis: a meta-analysis. *Am Heart J, 154*(5), 830-837. doi: 10.1016/j.ahj.2007.06.037

Beagle, B., Yang, T. L., Hung, J., Cogger, E. A., Moriarty, D. J., & Caudill, M. A. (2005). The glycine N-methyltransferase (GNMT) 1289 C->T variant influences plasma total homocysteine concentrations in young women after restricting folate intake. *J Nutr, 135*(12), 2780-2785.

Benowitz, N. L. (2003). Cigarette smoking and cardiovascular disease: pathophysiology and implications for treatment. *Prog Cardiovasc Dis, 46*(1), 91-111.

Benowitz, N. L., Jacob, P., 3rd, Jones, R. T., & Rosenberg, J. (1982). Interindividual variability in the metabolism and cardiovascular effects of nicotine in man. *J Pharmacol Exp Ther, 221*(2), 368-372.

Bermudez, E. A., Rifai, N., Buring, J. E., Manson, J. E., & Ridker, P. M. (2002). Relation between markers of systemic vascular inflammation and smoking in women. *Am J Cardiol, 89*(9), 1117-1119.

Biernacka, J. M., Tang, R., Li, J., McDonnell, S. K., Rabe, K. G., Sinnwell, J. P., . . . Fridley, B. L. (2009). Assessment of genotype imputation methods. *BMC Proc, 3 Suppl 7*, S5.

Board, P. G. (1981). Biochemical genetics of glutathione-S-transferase in man. *Am J Hum Genet, 33*(1), 36-43.

Boushey, C. J., Beresford, S. A., Omenn, G. S., & Motulsky, A. G. (1995). A quantitative assessment of plasma homocysteine as a risk factor for vascular disease. Probable benefits of increasing folic acid intakes. *JAMA, 274*(13), 1049-1057.

Brattstrom, L., Lindgren, A., Israelsson, B., Andersson, A., & Hultberg, B. (1994). Homocysteine and cysteine: determinants of plasma levels in middle-aged and elderly subjects. *J Intern Med, 236*(6), 633-641.

Braunwald, E. (1997). Shattuck lecture--cardiovascular medicine at the turn of the millennium: triumphs, concerns, and opportunities. *N Engl J Med, 337*(19), 1360-1369. doi: 10.1056/NEJM199711063371906

Breslow, J. L. (1997). Cardiovascular disease burden increases, NIH funding decreases. *Nat Med, 3*(6), 600-601.

Brook, R. D., Rajagopalan, S., Pope, C. A., 3rd, Brook, J. R., Bhatnagar, A., Diez-Roux, A. V., . . . Metabolism. (2010). Particulate matter air pollution and cardiovascular disease: An update to the scientific statement from the American Heart Association. *Circulation, 121*(21), 2331-2378. doi: 10.1161/CIR.0b013e3181dbece1

Browning, B. L., & Browning, S. R. (2009). A unified approach to genotype imputation and haplotype-phase inference for large data sets of trios and unrelated individuals. *Am J Hum Genet, 84*(2), 210-223. doi: 10.1016/j.ajhg.2009.01.005

Brustolin, S., Giugliani, R., & Felix, T. M. (2010). Genetics of homocysteine metabolism and associated disorders. *Braz J Med Biol Res, 43*(1), 1-7.

Buijs, R. M., Scheer, F. A., Kreier, F., Yi, C., Bos, N., Goncharuk, V. D., & Kalsbeek, A. (2006). Organization of circadian functions: interaction with the body. *Prog.Brain Res., 153*, 341-360. doi: S0079-6123(06)53020-1 [pii];10.1016/S0079-6123(06)53020-1 [doi]

Bunger, M. K., Wilsbacher, L. D., Moran, S. M., Clendenin, C., Radcliffe, L. A., Hogenesch, J. B., . . . Bradfield, C. A. (2000). Mop3 is an essential component of the master circadian pacemaker in mammals. *Cell, 103*(7), 1009-1017. doi: S0092-8674(00)00205-1 [pii]

Burgner, D., Davila, S., Breunis, W. B., Ng, S. B., Li, Y., Bonnard, C., . . . International Kawasaki Disease Genetics, C. (2009). A genome-wide association study identifies novel and functionally related susceptibility Loci for Kawasaki disease. *PLoS Genet, 5*(1), e1000319. doi: 10.1371/journal.pgen.1000319

Burns, D. M. (2003). Epidemiology of smoking-induced cardiovascular disease. *Prog Cardiovasc Dis, 46*(1), 11-29.

Bush, W. S., & Moore, J. H. (2012). Chapter 11: Genome-wide association studies. *PLoS Comput Biol, 8*(12), e1002822. doi: 10.1371/journal.pcbi.1002822

Byrne, J. A., Grieve, D. J., Bendall, J. K., Li, J. M., Gove, C., Lambeth, J. D., . . . Shah, A. M. (2003). Contrasting roles of NADPH oxidase isoforms in pressure-overload versus angiotensin II-induced cardiac hypertrophy. *Circ Res, 93*(9), 802-805. doi: 10.1161/01.RES.0000099504.30207.F5

Cai, H., & Harrison, D. G. (2000). Endothelial dysfunction in cardiovascular diseases: the role of oxidant stress. *Circ Res, 87*(10), 840-844.

Cao, J. J., Thach, C., Manolio, T. A., Psaty, B. M., Kuller, L. H., Chaves, P. H., . . . Cushman, M. (2003). C-reactive protein, carotid intima-media thickness, and incidence of ischemic stroke in the elderly: the Cardiovascular Health Study. *Circulation, 108*(2), 166-170. doi: 10.1161/01.CIR.0000079160.07364.6A

Carlson, C. S., Aldred, S. F., Lee, P. K., Tracy, R. P., Schwartz, S. M., Rieder, M., . . . Reiner, A. P. (2005). Polymorphisms within the C-reactive protein (CRP) promoter region are associated with plasma CRP levels. *Am.J.Hum.Genet., 77*(1), 64-77. doi: S0002-9297(07)60902-1 [pii];10.1086/431366 [doi]

Cattaneo, M. (1999). Hyperhomocysteinemia, atherosclerosis and thrombosis. *Thromb Haemost, 81*(2), 165-176.

Caudill, M. A., Dellschaft, N., Solis, C., Hinkis, S., Ivanov, A. A., Nash-Barboza, S., . . . Vermeylen, F. (2009). Choline intake, plasma riboflavin, and the phosphatidylethanolamine N-methyltransferase G5465A genotype predict plasma homocysteine in folate-deplete Mexican-American men with the methylenetetrahydrofolate reductase 677TT genotype. *J Nutr, 139*(4), 727-733. doi: 10.3945/jn.108.100222

Chen, F., Haigh, S., Barman, S., & Fulton, D. J. (2012). From form to function: the role of Nox4 in the cardiovascular system. *Front Physiol, 3*, 412. doi: 10.3389/fphys.2012.00412

Chen, M. H., & Yang, Q. (2010). GWAF: an R package for genome-wide association analyses with family data. *Bioinformatics, 26*(4), 580-581. doi: 10.1093/bioinformatics/btp710

Chen, Z., & Boreham, J. (2002). Smoking and cardiovascular disease. *Semin Vasc Med, 2*(3), 243-252. doi: 10.1055/s-2002-35392

Chenevix-Trench, G., Young, J., Coggan, M., & Board, P. (1995). Glutathione S-transferase M1 and T1 polymorphisms: susceptibility to colon cancer and age of onset. *Carcinogenesis, 16*(7), 1655-1657.

Clarkson, T. B., Weingand, K. W., Kaplan, J. R., & Adams, M. R. (1987). Mechanisms of atherogenesis. *Circulation, 76*(1 Pt 2), I20-28.

Clayton, D. G., Walker, N. M., Smyth, D. J., Pask, R., Cooper, J. D., Maier, L. M., . . . Todd, J. A. (2005). Population structure, differential bias and genomic control in a large-scale, case-control association study. *Nat Genet, 37*(11), 1243-1246. doi: 10.1038/ng1653

Clyne, B., & Olshaker, J. S. (1999). The C-reactive protein. *J Emerg Med, 17*(6), 1019-1025.

Coburn, C. T., Knapp, F. F., Jr., Febbraio, M., Beets, A. L., Silverstein, R. L., & Abumrad, N. A. (2000). Defective uptake and utilization of long chain fatty acids in muscle and adipose tissues of CD36 knockout mice. *J.Biol.Chem., 275*(42), 32523-32529. doi: 10.1074/jbc.M003826200 [doi];M003826200 [pii]

Collinsova, M., Strakova, J., Jiracek, J., & Garrow, T. A. (2006). Inhibition of betaine-homocysteine S-methyltransferase causes hyperhomocysteinemia in mice. *J Nutr, 136*(6), 1493-1497.

Conen, D., & Ridker, P. M. (2007). Clinical significance of high-sensitivity C-reactive protein in cardiovascular disease. *Biomark.Med., 1*(2), 229-241. doi: 10.2217/17520363.1.2.229 [doi]

Conklin, D. J., & Bhatnagar, A. (2011). Are glutathione S-transferase null genotypes "null and void" of risk for ischemic vascular disease? *Circ Cardiovasc Genet, 4*(4), 339-341. doi: 10.1161/CIRCGENETICS.111.960526

Conklin, D. J., Haberzettl, P., Lesgards, J. F., Prough, R. A., Srivastava, S., & Bhatnagar, A. (2009). Increased sensitivity of glutathione S-transferase P-null mice to cyclophosphamide-induced urinary bladder toxicity. *J Pharmacol Exp Ther, 331*(2), 456-469. doi: 10.1124/jpet.109.156513

Conrad, D. F., Andrews, T. D., Carter, N. P., Hurles, M. E., & Pritchard, J. K. (2006). A high-resolution survey of deletion polymorphism in the human genome. *Nat Genet, 38*(1), 75-81. doi: 10.1038/ng1697

Coon, H., Leppert, M. F., Eckfeldt, J. H., Oberman, A., Myers, R. H., Peacock, J. M., . . . Heiss, G. (2001). Genome-wide linkage analysis of lipids in the Hypertension Genetic Epidemiology Network (HyperGEN) Blood Pressure Study. *Arterioscler.Thromb.Vasc.Biol., 21*(12), 1969-1976.

Crawford, D. C., Sanders, C. L., Qin, X., Smith, J. D., Shephard, C., Wong, M., . . . Nickerson, D. A. (2006). Genetic variation is associated with C-reactive protein levels in the Third National Health and Nutrition Examination Survey. *Circulation, 114*(23), 2458-2465. doi: CIRCULATIONAHA.106.615740 [pii];10.1161/CIRCULATIONAHA.106.615740 [doi]

Cresci, M., Foffa, I., Ait-Ali, L., Pulignani, S., Gianicolo, E. A., Botto, N., . . . Andreassi, M. G. (2011). Maternal and paternal environmental risk factors, metabolizing GSTM1 and GSTT1 polymorphisms, and congenital heart disease. *Am J Cardiol, 108*(11), 1625-1631. doi: 10.1016/j.amjcard.2011.07.022

Curtis, A. M., Cheng, Y., Kapoor, S., Reilly, D., Price, T. S., & Fitzgerald, G. A. (2007). Circadian variation of blood pressure and the vascular response to asynchronous stress. *Proc.Natl.Acad.Sci.U.S.A, 104*(9), 3450-3455. doi: 0611680104 [pii];10.1073/pnas.0611680104 [doi]

Danenberg, H. D., Szalai, A. J., Swaminathan, R. V., Peng, L., Chen, Z., Seifert, P., . . . Edelman, E. R. (2003). Increased thrombosis after arterial injury in human C-reactive protein-transgenic mice. *Circulation, 108*(5), 512-515. doi: 10.1161/01.CIR.0000085568.13915.1E [doi];01.CIR.0000085568.13915.1E [pii]

Danesh, J., Wheeler, J. G., Hirschfield, G. M., Eda, S., Eiriksdottir, G., Rumley, A., . . . Gudnason, V. (2004a). C-reactive protein and other circulating markers of inflammation in the prediction of coronary heart disease. *N.Engl.J.Med., 350*(14), 1387-1397. doi: 10.1056/NEJMoa032804 [doi];350/14/1387 [pii]

Danesh, J., Wheeler, J. G., Hirschfield, G. M., Eda, S., Eiriksdottir, G., Rumley, A., . . . Gudnason, V. (2004b). C-reactive protein and other circulating markers of inflammation in the prediction of coronary heart disease. *N Engl J Med, 350*(14), 1387-1397. doi: 10.1056/NEJMoa032804

Das, T., Sen, A. K., Kempf, T., Pramanik, S. R., Mandal, C., & Mandal, C. (2003). Induction of glycosylation in human C-reactive protein under different pathological conditions. *Biochem J, 373*(Pt 2), 345-355. doi: 10.1042/BJ20021701

Davies, M. J. (1990). A macro and micro view of coronary vascular insult in ischemic heart disease. *Circulation, 82*(3 Suppl), II38-46.

Dehghan, A., Dupuis, J., Barbalic, M., Bis, J. C., Eiriksdottir, G., Lu, C., . . . Chasman, D. I. (2011). Meta-analysis of genome-wide association studies in >80 000 subjects identifies multiple loci for C-reactive protein levels. *Circulation, 123*(7), 731-738. doi: CIRCULATIONAHA.110.948570 [pii];10.1161/CIRCULATIONAHA.110.948570 [doi]

Design of the Women's Health Initiative clinical trial and observational study. The Women's Health Initiative Study Group. (1998). *Control Clin.Trials, 19*(1), 61-109. doi: S0197245697000780 [pii]

Devlin, B., & Risch, N. (1995). A comparison of linkage disequilibrium measures for fine-scale mapping. *Genomics, 29*(2), 311-322. doi: 10.1006/geno.1995.9003

Diaz, M. N., Frei, B., Vita, J. A., & Keaney, J. F., Jr. (1997). Antioxidants and atherosclerotic heart disease. *N Engl J Med, 337*(6), 408-416. doi: 10.1056/NEJM199708073370607

Doll, R., & Peto, R. (1978). Cigarette smoking and bronchial carcinoma: dose and time relationships among regular smokers and lifelong non-smokers. *J Epidemiol Community Health, 32*(4), 303-313.

Doney, A. S., Lee, S., Leese, G. P., Morris, A. D., & Palmer, C. N. (2005). Increased cardiovascular morbidity and mortality in type 2 diabetes is associated with the glutathione S transferase theta-null genotype: a Go-DARTS study. *Circulation, 111*(22), 2927-2934. doi: 10.1161/CIRCULATIONAHA.104.509224

Doumatey, A. P., Chen, G., Tekola, A. F., Zhou, J., Erdos, M., Shriner, D., . . . Rotimi, C. N. (2012). C-reactive protein (CRP) promoter polymorphisms influence circulating CRP levels in a genome-wide association study of African Americans. *Hum.Mol.Genet.* doi: dds133 [pii];10.1093/hmg/dds133 [doi]

Drysdale, C. M., McGraw, D. W., Stack, C. B., Stephens, J. C., Judson, R. S., Nandabalan, K., . . . Liggett, S. B. (2000). Complex promoter and coding region beta 2-adrenergic receptor haplotypes alter receptor expression and predict in vivo responsiveness. *Proc Natl Acad Sci U S A, 97*(19), 10483-10488.

Edberg, J. C., Wu, J., Langefeld, C. D., Brown, E. E., Marion, M. C., McGwin, G., Jr., . . . Kimberly, R. P. (2008). Genetic variation in the CRP promoter: association with systemic lupus erythematosus. *Hum.Mol.Genet., 17*(8), 1147-1155. doi: ddn004 [pii];10.1093/hmg/ddn004 [doi]

Esfahani, S. T., Cogger, E. A., & Caudill, M. A. (2003). Heterogeneity in the prevalence of methylenetetrahydrofolate reductase gene polymorphisms in women of different ethnic groups. *J Am Diet Assoc, 103*(2), 200-207. doi: 10.1053/jada.2003.50030

Falcone, D. J., McCaffrey, T. A., & Vergilio, J. A. (1991). Stimulation of macrophage urokinase expression by polyanions is protein kinase C-dependent and requires protein and RNA synthesis. *J Biol Chem, 266*(33), 22726-22732.

Falk, E. (1996). [Plaque vulnerability and disruption]. *Rev Clin Esp, 196*(4 Monografico), 6-12.

Fallin, D., & Schork, N. J. (2000). Accuracy of haplotype frequency estimation for biallelic loci, via the expectation-maximization algorithm for unphased diploid genotype data. *Am J Hum Genet, 67*(4), 947-959. doi: 10.1086/303069

Falush, D., Stephens, M., & Pritchard, J. K. (2007). Inference of population structure using multilocus genotype data: dominant markers and null alleles. *Mol Ecol Notes, 7*(4), 574-578. doi: 10.1111/j.1471-8286.2007.01758.x

Fox, E. R., Benjamin, E. J., Sarpong, D. F., Rotimi, C. N., Wilson, J. G., Steffes, M. W., . . . Taylor, H. A., Jr. (2008). Epidemiology, heritability, and genetic linkage of C-reactive protein in African Americans (from the Jackson Heart Study). *Am.J.Cardiol., 102*(7), 835-841. doi: S0002-9149(08)00913-2 [pii];10.1016/j.amjcard.2008.05.049 [doi]

Frosst, P., Blom, H. J., Milos, R., Goyette, P., Sheppard, C. A., Matthews, R. G., . . . et al. (1995). A candidate genetic risk factor for vascular disease: a common mutation in methylenetetrahydrofolate reductase. *Nat Genet, 10*(1), 111-113. doi: 10.1038/ng0595-111

Fry, A. E., Ghansa, A., Small, K. S., Palma, A., Auburn, S., Diakite, M., . . . Kwiatkowski, D. P. (2009). Positive selection of a CD36 nonsense variant in sub-Saharan Africa, but no association with severe malaria phenotypes. *Hum.Mol.Genet., 18*(14), 2683-2692. doi: ddp192 [pii];10.1093/hmg/ddp192 [doi]

Gabay, C. (2006). Interleukin-6 and chronic inflammation. *Arthritis Res.Ther., 8 Suppl 2*, S3. doi: ar1917 [pii];10.1186/ar1917 [doi]

Garlanda, C., Bottazzi, B., Bastone, A., & Mantovani, A. (2005). Pentraxins at the crossroads between innate immunity, inflammation, matrix deposition, and female fertility. *Annu Rev Immunol, 23*, 337-366. doi: 10.1146/annurev.immunol.23.021704.115756

Garrod, A. E. (1996). The incidence of alkaptonuria: a study in chemical individuality. 1902. *Mol Med, 2*(3), 274-282.

Gaustadnes, M., Rudiger, N., Rasmussen, K., & Ingerslev, J. (2000). Intermediate and severe hyperhomocysteinemia with thrombosis: a study of genetic determinants. *Thromb Haemost, 83*(4), 554-558.

Geiszt, M., Kopp, J. B., Varnai, P., & Leto, T. L. (2000). Identification of renox, an NAD(P)H oxidase in kidney. *Proc Natl Acad Sci U S A, 97*(14), 8010-8014. doi: 10.1073/pnas.130135897

Genomes Project, C., Abecasis, G. R., Altshuler, D., Auton, A., Brooks, L. D., Durbin, R. M., . . . McVean, G. A. (2010). A map of human genome variation from population-scale sequencing. *Nature, 467*(7319), 1061-1073. doi: 10.1038/nature09534

Ghosh, A., Murugesan, G., Chen, K., Zhang, L., Wang, Q., Febbraio, M., . . . Silverstein, R. L. (2011). Platelet CD36 surface expression levels affect functional responses to oxidized LDL and are associated with inheritance of specific genetic polymorphisms. *Blood, 117*(23), 6355-6366. doi: 10.1182/blood-2011-02-338582

Glantz, S. A., & Parmley, W. W. (1991). Passive smoking and heart disease. Epidemiology, physiology, and biochemistry. *Circulation, 83*(1), 1-12.

Goyal, T., Mitra, S., Khaidakov, M., Wang, X., Singla, S., Ding, Z., . . . Mehta, J. L. (2012). Current Concepts of the Role of Oxidized LDL Receptors in Atherosclerosis. *Curr.Atheroscler.Rep.* doi: 10.1007/s11883-012-0228-1 [doi]

Greenfield, J. R., Samaras, K., Jenkins, A. B., Kelly, P. J., Spector, T. D., Gallimore, J. R., . . . Campbell, L. V. (2004). Obesity is an important determinant of baseline serum C-reactive protein concentration in monozygotic twins, independent of genetic influences. *Circulation, 109*(24), 3022-3028. doi: 10.1161/01.CIR.0000130640.77501.79 [doi];01.CIR.0000130640.77501.79 [pii]

Griendling, K. K., & Alexander, R. W. (1997). Oxidative stress and cardiovascular disease. *Circulation, 96*(10), 3264-3265.

Griffith, O. L., Montgomery, S. B., Bernier, B., Chu, B., Kasaian, K., Aerts, S., . . . Open Regulatory Annotation, C. (2008). ORegAnno: an open-access community-driven resource for regulatory annotation. *Nucleic Acids Res, 36*(Database issue), D107-113. doi: 10.1093/nar/gkm967

Guan, Y., & Stephens, M. (2008). Practical issues in imputation-based association mapping. *PLoS Genet, 4*(12), e1000279. doi: 10.1371/journal.pgen.1000279

Gudnason, V., Stansbie, D., Scott, J., Bowron, A., Nicaud, V., & Humphries, S. (1998). C677T (thermolabile alanine/valine) polymorphism in methylenetetrahydrofolate reductase (MTHFR): its frequency and impact on plasma homocysteine concentration in different European populations. EARS group. *Atherosclerosis, 136*(2), 347-354.

Guzik, T. J., Sadowski, J., Guzik, B., Jopek, A., Kapelak, B., Przybylowski, P., . . . Channon, K. M. (2006). Coronary artery superoxide production and nox isoform expression in human coronary artery disease. *Arterioscler Thromb Vasc Biol, 26*(2), 333-339. doi: 10.1161/01.ATV.0000196651.64776.51

Habdous, M., Siest, G., Herbeth, B., Vincent-Viry, M., & Visvikis, S. (2004). [Glutathione S-transferases genetic polymorphisms and human diseases: overview of epidemiological studies]. *Ann Biol Clin (Paris), 62*(1), 15-24.

Hackam, D. G., & Anand, S. S. (2003). Emerging risk factors for atherosclerotic vascular disease: a critical review of the evidence. *JAMA, 290*(7), 932-940. doi: 10.1001/jama.290.7.932

Han, J., Hajjar, D. P., Febbraio, M., & Nicholson, A. C. (1997). Native and modified low density lipoproteins increase the functional expression of the macrophage class B scavenger receptor, CD36. *J Biol Chem, 272*(34), 21654-21659.

Handberg, A., Levin, K., Hojlund, K., & Beck-Nielsen, H. (2006). Identification of the oxidized low-density lipoprotein scavenger receptor CD36 in plasma: a novel marker of insulin resistance. *Circulation, 114*(11), 1169-1176. doi: CIRCULATIONAHA.106.626135 [pii];10.1161/CIRCULATIONAHA.106.626135 [doi]

Hansson, G. K., & Libby, P. (2006). The immune response in atherosclerosis: a double-edged sword. *Nat Rev Immunol, 6*(7), 508-519. doi: 10.1038/nri1882

Harris, T. B., Ferrucci, L., Tracy, R. P., Corti, M. C., Wacholder, S., Ettinger, W. H., Jr., . . . Wallace, R. (1999). Associations of elevated interleukin-6 and C-reactive protein levels with mortality in the elderly. *Am.J.Med., 106*(5), 506-512. doi: S0002934399000662 [pii]

Haverkate, F., Thompson, S. G., Pyke, S. D., Gallimore, J. R., & Pepys, M. B. (1997). Production of C-reactive protein and risk of coronary events in stable and unstable angina. European Concerted Action on Thrombosis and Disabilities Angina Pectoris Study Group. *Lancet, 349*(9050), 462-466.

Hayes, J. D., Flanagan, J. U., & Jowsey, I. R. (2005). Glutathione transferases. *Annu Rev Pharmacol Toxicol, 45*, 51-88. doi: 10.1146/annurev.pharmtox.45.120403.095857

Hayes, J. D., & Strange, R. C. (2000). Glutathione S-transferase polymorphisms and their biological consequences. *Pharmacology, 61*(3), 154-166. doi: 28396

Hazra, A., Kraft, P., Lazarus, R., Chen, C., Chanock, S. J., Jacques, P., . . . Hunter, D. J. (2009). Genome-wide significant predictors of metabolites in the one-carbon metabolism pathway. *Hum Mol Genet, 18*(23), 4677-4687. doi: 10.1093/hmg/ddp428

Heidenreich, P. A., Trogdon, J. G., Khavjou, O. A., Butler, J., Dracup, K., Ezekowitz, M. D., . . . Outcomes, R. (2011). Forecasting the future of cardiovascular disease in the United States: a policy statement from the American Heart Association. *Circulation, 123*(8), 933-944. doi: 10.1161/CIR.0b013e31820a55f5

Heijmans, B. T., Gussekloo, J., Kluft, C., Droog, S., Lagaay, A. M., Knook, D. L., . . . Slagboom, E. P. (1999). Mortality risk in men is associated with a common

mutation in the methylene-tetrahydrofolate reductase gene (MTHFR). *Eur J Hum Genet, 7*(2), 197-204. doi: 10.1038/sj.ejhg.5200283

Hellenbrand, W., Seidler, A., Robra, B. P., Vieregge, P., Oertel, W. H., Joerg, J., . . . Ulm, G. (1997). Smoking and Parkinson's disease: a case-control study in Germany. *Int J Epidemiol, 26*(2), 328-339.

Henry, O. R., Benghuzzi, H., Taylor, H. A., Jr., Tucci, M., Butler, K., & Jones, L. (2012). Suppression of homocysteine levels by vitamin B12 and folates: age and gender dependency in the Jackson Heart Study. *Am J Med Sci, 344*(2), 110-115. doi: 10.1097/MAJ.0b013e31823782a5

Hirschhorn, J. N., & Daly, M. J. (2005). Genome-wide association studies for common diseases and complex traits. *Nat Rev Genet, 6*(2), 95-108. doi: 10.1038/nrg1521

Homocysteine Studies, C. (2002). Homocysteine and risk of ischemic heart disease and stroke: a meta-analysis. *JAMA, 288*(16), 2015-2022.

Honein, M. A., Paulozzi, L. J., Mathews, T. J., Erickson, J. D., & Wong, L. Y. (2001). Impact of folic acid fortification of the US food supply on the occurrence of neural tube defects. *JAMA, 285*(23), 2981-2986.

Hong, K. W., Go, M. J., Jin, H. S., Lim, J. E., Lee, J. Y., Han, B. G., . . . Oh, B. (2010). Genetic variations in ATP2B1, CSK, ARSG and CSMD1 loci are related to blood pressure and/or hypertension in two Korean cohorts. *J Hum Hypertens, 24*(6), 367-372. doi: 10.1038/jhh.2009.86

Honma, S., Ikeda, M., Abe, H., Tanahashi, Y., Namihira, M., Honma, K., & Nomura, M. (1998). Circadian oscillation of BMAL1, a partner of a mammalian clock gene Clock, in rat suprachiasmatic nucleus. *Biochem.Biophys.Res.Commun., 250*(1), 83-87. doi: S0006291X98992750 [pii]

Howie, B. N., Donnelly, P., & Marchini, J. (2009). A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS.Genet., 5*(6), e1000529. doi: 10.1371/journal.pgen.1000529 [doi]

Howie, B. N., Donnelly, P., & Marchini, J. (2009). A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet, 5*(6), e1000529. doi: 10.1371/journal.pgen.1000529

Huedo-Medina, T. B., Sanchez-Meca, J., Marin-Martinez, F., & Botella, J. (2006). Assessing heterogeneity in meta-analysis: Q statistic or I2 index? *Psychol Methods, 11*(2), 193-206. doi: 10.1037/1082-989X.11.2.193

Hunter, D. J. (2005). Gene-environment interactions in human diseases. *Nat Rev Genet, 6*(4), 287-298. doi: 10.1038/nrg1578

Husemoen, L. L., Thomsen, T. F., Fenger, M., Jorgensen, H. L., & Jorgensen, T. (2003). Contribution of thermolabile methylenetetrahydrofolate reductase variant to total plasma homocysteine levels in healthy men and women. Inter99 (2). *Genet Epidemiol, 24*(4), 322-330. doi: 10.1002/gepi.10239

Ikeda, R., Ishii, K., Hoshikawa, Y., Azumi, J., Arakaki, Y., Yasui, T., . . . Shiota, G. (2011). Reactive oxygen species and NADPH oxidase 4 induced by transforming growth factor beta1 are the therapeutic targets of polyenylphosphatidylcholine in the suppression of human hepatic stellate cell activation. *Inflamm Res, 60*(6), 597-604. doi: 10.1007/s00011-011-0309-6

Indrayan, A., & Sarmukaddam, S. B. (2001). *Medical biostatistics*. New York: Marcel Dekker.

International HapMap, C. (2005). A haplotype map of the human genome. *Nature, 437*(7063), 1299-1320. doi: 10.1038/nature04226

Ioannidis, J. P. (2008). Calibration of credibility of agnostic genome-wide associations. *Am J Med Genet B Neuropsychiatr Genet, 147B*(6), 964-972. doi: 10.1002/ajmg.b.30721

Jacques, P. F., Bostom, A. G., Williams, R. R., Ellison, R. C., Eckfeldt, J. H., Rosenberg, I. H., . . . Rozen, R. (1996). Relation between folate status, a common mutation in methylenetetrahydrofolate reductase, and plasma homocysteine concentrations. *Circulation, 93*(1), 7-9.

Ji, S. R., Wu, Y., Potempa, L. A., Qiu, Q., & Zhao, J. (2006). Interactions of C-reactive protein with low-density lipoproteins: implications for an active role of modified C-reactive protein in atherosclerosis. *Int.J.Biochem.Cell Biol., 38*(4), 648-661. doi: S1357-2725(05)00384-5 [pii];10.1016/j.biocel.2005.11.004 [doi]

Jockel, K. H., Lehmann, N., Jaeger, B. R., Moebus, S., Mohlenkamp, S., Schmermund, A., . . . Erbel, R. (2009). Smoking cessation and subclinical atherosclerosis--results from the Heinz Nixdorf Recall Study. *Atherosclerosis, 203*(1), 221-227. doi: 10.1016/j.atherosclerosis.2008.05.041

Johnson, M. A., Hausman, D. B., Davey, A., Poon, L. W., Allen, R. H., Stabler, S. P., & Georgia Centenarian, S. (2010). Vitamin B12 deficiency in African American and white octogenarians and centenarians in Georgia. *J Nutr Health Aging, 14*(5), 339-345.

Jonasson, L., Holm, J., Skalli, O., Bondjers, G., & Hansson, G. K. (1986). Regional accumulations of T cells, macrophages, and smooth muscle cells in the human atherosclerotic plaque. *Arteriosclerosis, 6*(2), 131-138.

Jones, C. I., Bray, S., Garner, S. F., Stephens, J., de Bono, B., Angenent, W. G., . . . Bloodomics, C. (2009). A functional genomics approach reveals novel quantitative trait loci associated with platelet signaling pathways. *Blood, 114*(7), 1405-1416. doi: 10.1182/blood-2009-02-202614

Kalra, V. K., Ying, Y., Deemer, K., Natarajan, R., Nadler, J. L., & Coates, T. D. (1994). Mechanism of cigarette smoke condensate induced adhesion of human monocytes to cultured endothelial cells. *J Cell Physiol, 160*(1), 154-162. doi: 10.1002/jcp.1041600118

Kanani, P. M., Sinkey, C. A., Browning, R. L., Allaman, M., Knapp, H. R., & Haynes, W. G. (1999). Role of oxidant stress in endothelial dysfunction produced by experimental hyperhomocyst(e)inemia in humans. *Circulation, 100*(11), 1161-1168.

Kang, S. S. (1995). Critical points for determining moderate hyperhomocyst(e)inaemia. *Eur J Clin Invest, 25*(11), 806-808.

Karlsson, B., Knutsson, A., & Lindahl, B. (2001). Is there an association between shift work and having a metabolic syndrome? Results from a population based study of 27,485 people. *Occup.Environ.Med., 58*(11), 747-752.

Keating, B. J., Tischfield, S., Murray, S. S., Bhangale, T., Price, T. S., Glessner, J. T., . . . Fitzgerald, G. A. (2008). Concept, design and implementation of a cardiovascular gene-centric 50 k SNP array for large-scale genomic association studies. *PLoS One, 3*(10), e3583. doi: 10.1371/journal.pone.0003583

Khera, A., McGuire, D. K., Murphy, S. A., Stanek, H. G., Das, S. R., Vongpatanasin, W., . . . de Lemos, J. A. (2005). Race and gender differences in C-reactive protein levels. *J.Am.Coll.Cardiol., 46*(3), 464-469. doi: S0735-1097(05)01042-9 [pii];10.1016/j.jacc.2005.04.051 [doi]

Khoo, J. C., Miller, E., McLoughlin, P., & Steinberg, D. (1988). Enhanced macrophage uptake of low density lipoprotein after self-aggregation. *Arteriosclerosis, 8*(4), 348-358.

Khoo, J. C., Miller, E., Pio, F., Steinberg, D., & Witztum, J. L. (1992). Monoclonal antibodies against LDL further enhance macrophage uptake of LDL aggregates. *Arterioscler Thromb, 12*(11), 1258-1266.

Kim, I. K., Zhang, C. Y., Chiang, P. K., & Cantoni, G. L. (1983). S-Adenosylhomocysteine hydrolase from hamster liver: purification and kinetic properties. *Arch Biochem Biophys, 226*(1), 65-72.

Kim, J., Kim, Y., Kim, H. T., Kim, D. W., Ha, Y., Kim, J., . . . Song, K. (2009). TC1(C8orf4) is a novel endothelial inflammatory regulator enhancing NF-kappaB activity. *J Immunol, 183*(6), 3996-4002. doi: 10.4049/jimmunol.0900956

Kim, Y., Kim, J., Park, J., Bang, S., Jung, Y., Choe, J., . . . Lee, I. (2006). TC1(C8orf4) is upregulated by IL-1beta/TNF-alpha and enhances proliferation of human follicular dendritic cells. *FEBS Lett, 580*(14), 3519-3524. doi: 10.1016/j.febslet.2006.05.036

Kluijtmans, L. A., Young, I. S., Boreham, C. A., Murray, L., McMaster, D., McNulty, H., . . . Whitehead, A. S. (2003). Genetic and nutritional factors contributing to hyperhomocysteinemia in young adults. *Blood, 101*(7), 2483-2488. doi: 10.1182/blood.V101.7.2483

Knutsson, A., Akerstedt, T., Jonsson, B. G., & Orth-Gomer, K. (1986). Increased risk of ischaemic heart disease in shift workers. *Lancet, 2*(8498), 89-92.

Koenig, W., Lowel, H., Baumert, J., & Meisinger, C. (2004). C-reactive protein modulates risk prediction based on the Framingham Score: implications for future risk assessment: results from a large cohort study in southern Germany. *Circulation, 109*(11), 1349-1353. doi: 10.1161/01.CIR.0000120707.98922.E3 [doi];01.CIR.0000120707.98922.E3 [pii]

Koenig, W., Sund, M., Frohlich, M., Fischer, H. G., Lowel, H., Doring, A., . . . Pepys, M. B. (1999). C-Reactive protein, a sensitive marker of inflammation, predicts future risk of coronary heart disease in initially healthy middle-aged men: results from the MONICA (Monitoring Trends and Determinants in Cardiovascular Disease) Augsburg Cohort Study, 1984 to 1992. *Circulation, 99*(2), 237-242.

Koyama, H., Raines, E. W., Bornfeldt, K. E., Roberts, J. M., & Ross, R. (1996). Fibrillar collagen inhibits arterial smooth muscle proliferation through regulation of Cdk2 inhibitors. *Cell, 87*(6), 1069-1078.

Kraus, D. M., Elliott, G. S., Chute, H., Horan, T., Pfenninger, K. H., Sanford, S. D., . . . Holers, V. M. (2006). CSMD1 is a novel multiple domain complement-regulatory protein highly expressed in the central nervous system and epithelial tissues. *J Immunol, 176*(7), 4419-4430.

Kuklina, E. V., Tong, X., George, M. G., & Bansil, P. (2012). Epidemiology and prevention of stroke: a worldwide perspective. *Expert Rev Neurother, 12*(2), 199-208. doi: 10.1586/ern.11.99

Lake, S. L., Lyon, H., Tantisira, K., Silverman, E. K., Weiss, S. T., Laird, N. M., & Schaid, D. J. (2003). Estimation and tests of haplotype-environment interaction when linkage phase is ambiguous. *Hum Hered, 55*(1), 56-65. doi: 71811

Lalouel, J. M., & Rohrwasser, A. (2002). Power and replication in case-control studies. *Am J Hypertens, 15*(2 Pt 1), 201-205.

Lambeth, J. D., Kawahara, T., & Diebold, B. (2007). Regulation of Nox and Duox enzymatic activity and expression. *Free Radic Biol Med, 43*(3), 319-331. doi: 10.1016/j.freeradbiomed.2007.03.028

Landmark, K. (2001). [Smoking and coronary heart disease]. *Tidsskr Nor Laegeforen, 121*(14), 1710-1712.

Lange, L. A., Carlson, C. S., Hindorff, L. A., Lange, E. M., Walston, J., Durda, J. P., . . . Reiner, A. P. (2006). Association of polymorphisms in the CRP gene with circulating C-reactive protein levels and cardiovascular events. *JAMA, 296*(22), 2703-2711. doi: 10.1001/jama.296.22.2703

Lange, L. A., Croteau-Chonka, D. C., Marvelle, A. F., Qin, L., Gaulton, K. J., Kuzawa, C. W., . . . Mohlke, K. L. (2010). Genome-wide association study of homocysteine levels in Filipinos provides evidence for CPS1 in women and a stronger MTHFR effect in young adults. *Hum Mol Genet, 19*(10), 2050-2058. doi: 10.1093/hmg/ddq062

Larkin, E. K., Rosen, C. L., Kirchner, H. L., Storfer-Isser, A., Emancipator, J. L., Johnson, N. L., . . . Redline, S. (2005). Variation of C-reactive protein levels in adolescents: association with sleep-disordered breathing and sleep duration. *Circulation, 111*(15), 1978-1984. doi: 111/15/1978 [pii];10.1161/01.CIR.0000161819.76138.5E [doi]

Law, M. R., Morris, J. K., & Wald, N. J. (1997). Environmental tobacco smoke exposure and ischaemic heart disease: an evaluation of the evidence. *BMJ, 315*(7114), 973-980.

Lawlor, D. A., Harbord, R. M., Sterne, J. A., Timpson, N., & Davey, S. G. (2008). Mendelian randomization: using genes as instruments for making causal inferences in epidemiology. *Stat.Med., 27*(8), 1133-1163. doi: 10.1002/sim.3034 [doi]

Lee, S., Choi, K., Ahn, H., Song, K., Choe, J., & Lee, I. (2005). TuJ1 (class III beta-tubulin) expression suggests dynamic redistribution of follicular dendritic cells in lymphoid tissue. *Eur J Cell Biol, 84*(2-3), 453-459.

Lewontin, R. C. (1988). On measures of gametic disequilibrium. *Genetics, 120*(3), 849-852.

Li, M., Li, C., & Guan, W. (2008). Evaluation of coverage variation of SNP chips for genome-wide association studies. *Eur J Hum Genet, 16*(5), 635-643. doi: 10.1038/sj.ejhg.5202007

Li, Y., Willer, C., Sanna, S., & Abecasis, G. (2009). Genotype imputation. *Annu Rev Genomics Hum Genet, 10*, 387-406. doi: 10.1146/annurev.genom.9.081307.164242

Li, Y., Willer, C. J., Ding, J., Scheet, P., & Abecasis, G. R. (2010). MaCH: using sequence and genotype data to estimate haplotypes and unobserved genotypes. *Genet Epidemiol, 34*(8), 816-834. doi: 10.1002/gepi.20533

Li, Z., & Vance, D. E. (2008). Phosphatidylcholine and choline homeostasis. *J Lipid Res, 49*(6), 1187-1194. doi: 10.1194/jlr.R700019-JLR200

Libby, P., Ridker, P. M., & Maseri, A. (2002). Inflammation and atherosclerosis. *Circulation, 105*(9), 1135-1143.

Lin, D. Y., & Zeng, D. (2010). Meta-analysis of genome-wide association studies: no efficiency gain in using individual participant data. *Genet Epidemiol, 34*(1), 60-66. doi: 10.1002/gepi.20435

Long, A. D., & Langley, C. H. (1999). The power of association studies to detect the contribution of candidate genetic loci to variation in complex traits. *Genome Res, 9*(8), 720-731.

Love-Gregory, L., Sherva, R., Schappe, T., Qi, J. S., McCrea, J., Klein, S., . . . Abumrad, N. A. (2011). Common CD36 SNPs reduce protein expression and may contribute to a protective atherogenic profile. *Hum.Mol.Genet., 20*(1), 193-201. doi: ddq449 [pii];10.1093/hmg/ddq449 [doi]

Love-Gregory, L., Sherva, R., Sun, L., Wasson, J., Schappe, T., Doria, A., . . . Abumrad, N. A. (2008). Variants in the CD36 gene associate with the metabolic syndrome and high-density lipoprotein cholesterol. *Hum.Mol.Genet., 17*(11), 1695-1704. doi: ddn060 [pii];10.1093/hmg/ddn060 [doi]

Lowe, G. D., Yarnell, J. W., Rumley, A., Bainton, D., & Sweetnam, P. M. (2001). C-reactive protein, fibrin D-dimer, and incident ischemic heart disease in the Speedwell study: are inflammation and fibrin turnover linked in pathogenesis? *Arterioscler Thromb Vasc Biol, 21*(4), 603-610.

Lowel, H., Doring, A., Schneider, A., Heier, M., Thorand, B., & Meisinger, C. (2005). The MONICA Augsburg surveys--basis for prospective cohort studies. *Gesundheitswesen, 67 Suppl 1*, S13-S18. doi: 10.1055/s-2005-858234 [doi]

Lusis, A. J., Fogelman, A. M., & Fonarow, G. C. (2004). Genetic basis of atherosclerosis: part II: clinical implications. *Circulation, 110*(14), 2066-2071. doi: 10.1161/01.CIR.0000143098.98869.F8

Lusis, A. J., Mar, R., & Pajukanta, P. (2004). Genetics of atherosclerosis. *Annu Rev Genomics Hum Genet, 5*, 189-218. doi: 10.1146/annurev.genom.5.061903.175930

Maemura, K., Layne, M. D., Watanabe, M., Perrell, M. A., Nagai, R., & Lee, M. E. (2001). Molecular mechanisms of morning onset of myocardial infarction. *Ann.N.Y.Acad.Sci., 947*, 398-402.

Malarstig, A., Buil, A., Souto, J. C., Clarke, R., Blanco-Vaca, F., Fontcuberta, J., . . . Precocious Coronary Artery Disease, c. (2009). Identification of ZNF366 and PTPRD as novel determinants of plasma homocysteine in a family-based genome-wide association study. *Blood, 114*(7), 1417-1422. doi: 10.1182/blood-2009-04-215269

Mansoor, M. A., Bergmark, C., Svardal, A. M., Lonning, P. E., & Ueland, P. M. (1995). Redox status and protein binding of plasma homocysteine and other aminothiols in patients with early-onset peripheral vascular disease. Homocysteine and peripheral vascular disease. *Arterioscler Thromb Vasc Biol, 15*(2), 232-240.

Marchini, J., Cardon, L. R., Phillips, M. S., & Donnelly, P. (2004). The effects of human population structure on large genetic association studies. *Nat Genet, 36*(5), 512-517. doi: 10.1038/ng1337

Marott, S. C., Nordestgaard, B. G., Zacho, J., Friberg, J., Jensen, G. B., Tybjaerg-Hansen, A., & Benn, M. (2010). Does elevated C-reactive protein increase atrial fibrillation risk? A Mendelian randomization of 47,000 individuals from the general population. *J.Am.Coll.Cardiol., 56*(10), 789-795. doi: S0735-1097(10)02305-3 [pii];10.1016/j.jacc.2010.02.066 [doi]

Martyn, K. D., Frederick, L. M., von Loehneysen, K., Dinauer, M. C., & Knaus, U. G. (2006). Functional analysis of Nox4 reveals unique characteristics compared to other NADPH oxidases. *Cell Signal, 18*(1), 69-82. doi: 10.1016/j.cellsig.2005.03.023

Mazzone, A., Cusa, C., Mazzucchelli, I., Vezzoli, M., Ottini, E., Ghio, S., . . . Zuccaro, P. (2001). Cigarette smoking and hypertension influence nitric oxide release and plasma levels of adhesion molecules. *Clin Chem Lab Med, 39*(9), 822-826. doi: 10.1515/CCLM.2001.136

McCarthy, M. I. (2008). Casting a wider net for diabetes susceptibility genes. *Nat.Genet., 40*(9), 1039-1040. doi: 10.1038/ng0908-1039 [doi]

McCully, K. S. (1996). Homocysteine and vascular disease. *Nat Med, 2*(4), 386-389.

McCully, K. S., & Carvalho, A. C. (1987). Homocysteine thiolactone, N-homocysteine thiolactonyl retinamide, and platelet aggregation. *Res Commun Chem Pathol Pharmacol, 56*(3), 349-360.

McKeever, M. P., Weir, D. G., Molloy, A., & Scott, J. M. (1991). Betaine-homocysteine methyltransferase: organ distribution in man, pig and rat and subcellular distribution in the rat. *Clin Sci (Lond), 81*(4), 551-556.

McLaughlin, T., Abbasi, F., Lamendola, C., Liang, L., Reaven, G., Schaaf, P., & Reaven, P. (2002). Differentiation between obesity and insulin resistance in the association with C-reactive protein. *Circulation, 106*(23), 2908-2912.

Medford, R. M., Dagi, T. F., Rosenson, R. S., & Offermann, M. K. (2013). Biomarkers and sustainable innovation in cardiovascular drug development: lessons from near and far afield. *Curr Atheroscler Rep, 15*(5), 321. doi: 10.1007/s11883-013-0321-0

Mendall, M. A., Patel, P., Ballam, L., Strachan, D., & Northfield, T. C. (1996). C reactive protein and its relation to cardiovascular risk factors: a population based cross sectional study. *BMJ, 312*(7038), 1061-1065.

Mercurius, K. O., & Morla, A. O. (1998). Inhibition of vascular smooth muscle cell growth by inhibition of fibronectin matrix assembly. *Circ Res, 82*(5), 548-556.

Mezzano, D., Munoz, X., Martinez, C., Cuevas, A., Panes, O., Aranda, E., . . . Leighton, F. (1999). Vegetarians and cardiovascular risk factors: hemostasis, inflammatory markers and plasma homocysteine. *Thromb Haemost, 81*(6), 913-917.

Miller, E. A., Pankow, J. S., Millikan, R. C., Bray, M. S., Ballantyne, C. M., Bell, D. A., . . . Li, R. (2003). Glutathione-S-transferase genotypes, smoking, and their association with markers of inflammation, hemostasis, and endothelial function: the atherosclerosis risk in communities (ARIC) study. *Atherosclerosis, 171*(2), 265-272.

Miyaoka, K., Kuwasako, T., Hirano, K., Nozaki, S., Yamashita, S., & Matsuzawa, Y. (2001). CD36 deficiency associated with insulin resistance. *Lancet, 357*(9257), 686-687. doi: S0140673600041386 [pii]

Morel, D. W., Hessler, J. R., & Chisolm, G. M. (1983). Low density lipoprotein cytotoxicity induced by free radical peroxidation of lipid. *J Lipid Res, 24*(8), 1070-1076.

Morley, J. J., & Kushner, I. (1982). Serum C-reactive protein levels in disease. *Ann N Y Acad Sci, 389*, 406-418.

Moyer, A. M., Salavaggione, O. E., Hebbring, S. J., Moon, I., Hildebrandt, M. A., Eckloff, B. W., . . . Weinshilboum, R. M. (2007). Glutathione S-transferase T1 and M1: gene sequence variation and functional genomics. *Clin Cancer Res, 13*(23), 7207-7216. doi: 10.1158/1078-0432.CCR-07-0635

Mudd, S. H., Finkelstein, J. D., Irreverre, F., & Laster, L. (1965). Transsulfuration in mammals. Microassays and tissue distributions of three enzymes of the pathway. *J Biol Chem, 240*(11), 4382-4392.

Mudd, S. H., & Poole, J. R. (1975). Labile methyl balances for normal humans on various dietary regimens. *Metabolism, 24*(6), 721-735.

Mullineaux, P. M., Karpinski, S., Jimenez, A., Cleary, S. P., Robinson, C., & Creissen, G. P. (1998). Identification of cDNAS encoding plastid-targeted glutathione peroxidase. *Plant J, 13*(3), 375-379.

Musunuru, K., Romaine, S. P., Lettre, G., Wilson, J. G., Volcik, K. A., Tsai, M. Y., . . . Rader, D. J. (2012). Multi-ethnic analysis of lipid-associated loci: the NHLBI CARe project. *PLoS.One., 7*(5), e36473. doi: 10.1371/journal.pone.0036473 [doi];PONE-D-11-06421 [pii]

Naito, Y., Tsujino, T., Kawasaki, D., Okumura, T., Morimoto, S., Masai, M., . . . Iwasaki, T. (2003). Circadian gene expression of clock genes and plasminogen activator inhibitor-1 in heart and aorta of spontaneously hypertensive and Wistar-Kyoto rats. *J.Hypertens., 21*(6), 1107-1115. doi: 10.1097/01.hjh.0000059048.65882.e4 [doi]

Nalls, M. A., Couper, D. J., Tanaka, T., van Rooij, F. J., Chen, M. H., Smith, A. V., . . . Ganesh, S. K. (2011). Multiple loci are associated with white blood cell phenotypes. *PLoS.Genet., 7*(6), e1002113. doi: 10.1371/journal.pgen.1002113 [doi];PGENETICS-D-10-00114 [pii]

Napoli, C., D'Armiento, F. P., Mancini, F. P., Postiglione, A., Witztum, J. L., Palumbo, G., & Palinski, W. (1997). Fatty streak formation occurs in human fetal aortas and is greatly enhanced by maternal hypercholesterolemia. Intimal accumulation of low density lipoprotein and its oxidation precede monocyte recruitment into early atherosclerotic lesions. *J Clin Invest, 100*(11), 2680-2690. doi: 10.1172/JCI119813

Navab, M., Berliner, J. A., Watson, A. D., Hama, S. Y., Territo, M. C., Lusis, A. J., . . . Fogelman, A. M. (1996). The Yin and Yang of oxidation in the development of the fatty streak. A review based on the 1994 George Lyman Duff Memorial Lecture. *Arterioscler Thromb Vasc Biol, 16*(7), 831-842.

Nelen, W. L., Blom, H. J., Thomas, C. M., Steegers, E. A., Boers, G. H., & Eskes, T. K. (1998). Methylenetetrahydrofolate reductase polymorphism affects the change in homocysteine and folate concentrations resulting from low dose folic acid supplementation in women with unexplained recurrent miscarriages. *J Nutr, 128*(8), 1336-1341.

Nelson, H. H., Wiencke, J. K., Christiani, D. C., Cheng, T. J., Zuo, Z. F., Schwartz, B. S., . . . et al. (1995). Ethnic differences in the prevalence of the homozygous deleted genotype of glutathione S-transferase theta. *Carcinogenesis, 16*(5), 1243-1245.

Ntais, C., Polycarpou, A., & Ioannidis, J. P. (2005). Association of GSTM1, GSTT1, and GSTP1 gene polymorphisms with the risk of prostate cancer: a meta-analysis. *Cancer Epidemiol Biomarkers Prev, 14*(1), 176-181.

Olshan, A. F., Li, R., Pankow, J. S., Bray, M., Tyroler, H. A., Chambless, L. E., . . . Bell, D. A. (2003). Risk of atherosclerosis: interaction of smoking and glutathione S-transferase genes. *Epidemiology, 14*(3), 321-327.

Omi, K., Ohashi, J., Patarapotikul, J., Hananantachai, H., Naka, I., Looareesuwan, S., & Tokunaga, K. (2003). CD36 polymorphism is associated with protection from cerebral malaria. *Am.J.Hum.Genet., 72*(2), 364-374. doi: S0002-9297(07)60545-X [pii];10.1086/346091 [doi]

Opie, L. H., & Lecour, S. (2007). The red wine hypothesis: from concepts to protective signalling molecules. *Eur Heart J, 28*(14), 1683-1693. doi: 10.1093/eurheartj/ehm149

Paalani, M., Lee, J. W., Haddad, E., & Tonstad, S. (2011). Determinants of inflammatory markers in a bi-ethnic population. *Ethn.Dis., 21*(2), 142-149.

Pankow, J. S., Folsom, A. R., Cushman, M., Borecki, I. B., Hopkins, P. N., Eckfeldt, J. H., & Tracy, R. P. (2001). Familial and genetic determinants of systemic markers of inflammation: the NHLBI family heart study. *Atherosclerosis, 154*(3), 681-689. doi: S0021915000005864 [pii]

Pare, G., Chasman, D. I., Parker, A. N., Zee, R. R., Malarstig, A., Seedorf, U., . . . Ridker, P. M. (2009). Novel associations of CPS1, MUT, NOX4, and DPEP1 with plasma homocysteine in a healthy population: a genome-wide evaluation of 13 974 participants in the Women's Genome Health Study. *Circ Cardiovasc Genet, 2*(2), 142-150. doi: 10.1161/CIRCGENETICS.108.829804

Park, J., Jung, Y., Kim, J., Kim, K. Y., Ahn, S. G., Song, K., & Lee, I. (2007). TC1 (C8orf4) is upregulated by cellular stress and mediates heat shock response. *Biochem Biophys Res Commun, 360*(2), 447-452. doi: 10.1016/j.bbrc.2007.06.077

Pepys, M. B., & Hirschfield, G. M. (2003). C-reactive protein: a critical update. *J Clin Invest, 111*(12), 1805-1812. doi: 10.1172/JCI18921

Perou, C. M., Sorlie, T., Eisen, M. B., van de Rijn, M., Jeffrey, S. S., Rees, C. A., . . . Botstein, D. (2000). Molecular portraits of human breast tumours. *Nature, 406*(6797), 747-752. doi: 10.1038/35021093

Perry, G. H., Ben-Dor, A., Tsalenko, A., Sampas, N., Rodriguez-Revenga, L., Tran, C. W., . . . Lee, C. (2008). The fine-scale and complex architecture of human copy-number variation. *Am J Hum Genet, 82*(3), 685-695. doi: 10.1016/j.ajhg.2007.12.010

Pietila, K. O., Harmoinen, A. P., Jokiniitty, J., & Pasternack, A. I. (1996). Serum C-reactive protein concentration in acute myocardial infarction and its relationship to mortality during 24 months of follow-up in patients under thrombolytic treatment. *Eur Heart J, 17*(9), 1345-1349.

Pober, J. S., & Sessa, W. C. (2007). Evolving functions of endothelial cells in inflammation. *Nat Rev Immunol, 7*(10), 803-815. doi: 10.1038/nri2171

Poddar, R., Sivasubramanian, N., DiBello, P. M., Robinson, K., & Jacobsen, D. W. (2001). Homocysteine induces expression and secretion of monocyte chemoattractant protein-1 and interleukin-8 in human aortic endothelial cells: implications for vascular disease. *Circulation, 103*(22), 2717-2723.

Portman, M. A. (2001). Molecular clock mechanisms and circadian rhythms intrinsic to the heart. *Circ.Res., 89*(12), 1084-1086.

Potenza, M. A., Marasciulo, F. L., Tarquinio, M., Tiravanti, E., Colantuono, G., Federici, A., . . . Montagnani, M. (2007). EGCG, a green tea polyphenol, improves endothelial function and insulin sensitivity, reduces blood pressure, and protects against myocardial I/R injury in SHR. *Am J Physiol Endocrinol Metab, 292*(5), E1378-1387. doi: 10.1152/ajpendo.00698.2006

Prandoni, P., Bilora, F., Marchiori, A., Bernardi, E., Petrobelli, F., Lensing, A. W., . . . Girolami, A. (2003). An association between atherosclerosis and venous thrombosis. *N Engl J Med, 348*(15), 1435-1441. doi: 10.1056/NEJMoa022157

Price, A. L., Patterson, N. J., Plenge, R. M., Weinblatt, M. E., Shadick, N. A., & Reich, D. (2006). Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet, 38*(8), 904-909. doi: 10.1038/ng1847

Price, A. L., Patterson, N. J., Plenge, R. M., Weinblatt, M. E., Shadick, N. A., & Reich, D. (2006). Principal components analysis corrects for stratification in genome-wide association studies. *Nat.Genet., 38*(8), 904-909. doi: ng1847 [pii];10.1038/ng1847 [doi]

Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics, 155*(2), 945-959.

Pulit, S. L., Voight, B. F., & de Bakker, P. I. (2010). Multiethnic genetic association studies improve power for locus discovery. *PLoS.One., 5*(9), e12600. doi: 10.1371/journal.pone.0012600 [doi]

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A., Bender, D., . . . Sham, P. C. (2007a). PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet, 81*(3), 559-575. doi: 10.1086/519795

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A., Bender, D., . . . Sham, P. C. (2007b). PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am.J.Hum.Genet., 81*(3), 559-575. doi: S0002-9297(07)61352-4 [pii];10.1086/519795 [doi]

Qayyum, R., Snively, B. M., Ziv, E., Nalls, M. A., Liu, Y., Tang, W., . . . Reiner, A. P. (2012). A meta-analysis and genome-wide association study of platelet count and mean platelet volume in african americans. *PLoS Genet, 8*(3), e1002491. doi: 10.1371/journal.pgen.1002491

Rader, D. J., & Daugherty, A. (2008). Translating molecular discoveries into new therapies for atherosclerosis. *Nature, 451*(7181), 904-913. doi: 10.1038/nature06796

Raimondi, S., Paracchini, V., Autrup, H., Barros-Dios, J. M., Benhamou, S., Boffetta, P., . . . Taioli, E. (2006). Meta- and pooled analysis of GSTT1 and lung cancer: a HuGE-GSEC review. *Am J Epidemiol, 164*(11), 1027-1042. doi: 10.1093/aje/kwj321

Ray, P., Le, M. Y., Riou, B., & Houle, T. T. (2010). Statistical evaluation of a biomarker. *Anesthesiology, 112*(4), 1023-1040. doi: 10.1097/ALN.0b013e3181d47604 [doi]

Rebbeck, T. R. (1997). Molecular epidemiology of the human glutathione S-transferase genotypes GSTM1 and GSTT1 in cancer susceptibility. *Cancer Epidemiol.Biomarkers Prev., 6*(9), 733-743.

Rebbeck, T. R., Spitz, M., & Wu, X. (2004). Assessing the function of genetic variants in candidate gene association studies. *Nat Rev Genet, 5*(8), 589-597. doi: 10.1038/nrg1403

Reich, D. E., & Lander, E. S. (2001). On the allelic spectrum of human disease. *Trends Genet, 17*(9), 502-510.

Reiner, A. P., Barber, M. J., Guan, Y., Ridker, P. M., Lange, L. A., Chasman, D. I., . . . Krauss, R. M. (2008). Polymorphisms of the HNF1A gene encoding hepatocyte nuclear factor-1 alpha are associated with C-reactive protein. *Am.J.Hum.Genet., 82*(5), 1193-1201. doi: S0002-9297(08)00227-9 [pii];10.1016/j.ajhg.2008.03.017 [doi]

Reiner, A. P., Beleza, S., Franceschini, N., Auer, P. L., Robinson, J. G., Kooperberg, C., . . . Tang, H. (2012). Genome-wide association and population genetic analysis of C-reactive protein in African American and Hispanic American women. *Am J Hum Genet, 91*(3), 502-512. doi: 10.1016/j.ajhg.2012.07.023

Ridker, P. M. (2010). Inflammation, high-sensitivity C-reactive protein, and vascular protection. *Tex.Heart Inst.J., 37*(1), 40-41.

Ridker, P. M., Buring, J. E., Rifai, N., & Cook, N. R. (2007). Development and validation of improved algorithms for the assessment of global cardiovascular risk in women: the Reynolds Risk Score. *JAMA, 297*(6), 611-619. doi: 297/6/611 [pii];10.1001/jama.297.6.611 [doi]

Ridker, P. M., Pare, G., Parker, A., Zee, R. Y., Danik, J. S., Buring, J. E., . . . Chasman, D. I. (2008). Loci related to metabolic-syndrome pathways including LEPR,HNF1A, IL6R, and GCKR associate with plasma C-reactive protein: the Women's Genome Health Study. *Am.J.Hum.Genet., 82*(5), 1185-1192. doi: S0002-9297(08)00225-5 [pii];10.1016/j.ajhg.2008.03.015 [doi]

Ridker, P. M., Rifai, N., Rose, L., Buring, J. E., & Cook, N. R. (2002). Comparison of C-reactive protein and low-density lipoprotein cholesterol levels in the prediction of first cardiovascular events. *N.Engl.J.Med., 347*(20), 1557-1565. doi: 10.1056/NEJMoa021993 [doi];347/20/1557 [pii]

Ridker, P. M., & Silvertown, J. D. (2008). Inflammation, C-reactive protein, and atherothrombosis. *J Periodontol, 79*(8 Suppl), 1544-1551. doi: 10.1902/jop.2008.080249

Risch, N., & Merikangas, K. (1996). The future of genetic studies of complex human diseases. *Science, 273*(5281), 1516-1517.

Risch, N. J. (2000). Searching for genetic determinants in the new millennium. *Nature, 405*(6788), 847-856. doi: 10.1038/35015718

Ritchie, M. D., Denny, J. C., Crawford, D. C., Ramirez, A. H., Weiner, J. B., Pulley, J. M., . . . Roden, D. M. (2010). Robust replication of genotype-phenotype associations across multiple diseases in an electronic medical record. *Am J Hum Genet, 86*(4), 560-572. doi: 10.1016/j.ajhg.2010.03.003

Robbie, L., & Libby, P. (2001). Inflammation and atherothrombosis. *Ann.N.Y.Acad.Sci., 947*, 167-179.

Roest, M., van der Schouw, Y. T., Grobbee, D. E., Tempelman, M. J., de Groot, P. G., Sixma, J. J., & Banga, J. D. (2001). Methylenetetrahydrofolate reductase 677 C/T genotype and cardiovascular disease mortality in postmenopausal women. *Am J Epidemiol, 153*(7), 673-679.

Roger, V. L., Go, A. S., Lloyd-Jones, D. M., Adams, R. J., Berry, J. D., Brown, T. M., . . . Wylie-Rosett, J. (2011). Heart disease and stroke statistics--2011 update: a report from the American Heart Association. *Circulation, 123*(4), e18-e209. doi: CIR.0b013e3182009701 [pii];10.1161/CIR.0b013e3182009701 [doi]

Rosenberg, N. A., Huang, L., Jewett, E. M., Szpiech, Z. A., Jankovic, I., & Boehnke, M. (2010). Genome-wide association studies in diverse populations. *Nat.Rev.Genet., 11*(5), 356-366. doi: nrg2760 [pii];10.1038/nrg2760 [doi]

Rosenberg, N. A., Mahajan, S., Ramachandran, S., Zhao, C., Pritchard, J. K., & Feldman, M. W. (2005). Clines, clusters, and the effect of study design on the inference of human population structure. *PLoS Genet, 1*(6), e70. doi: 10.1371/journal.pgen.0010070

Rosenberg, N. A., Pritchard, J. K., Weber, J. L., Cann, H. M., Kidd, K. K., Zhivotovsky, L. A., & Feldman, M. W. (2002). Genetic structure of human populations. *Science, 298*(5602), 2381-2385. doi: 10.1126/science.1078311

Ross, R. (1986). The pathogenesis of atherosclerosis--an update. *N Engl J Med, 314*(8), 488-500. doi: 10.1056/NEJM198602203140806

Ross, R. (1993). The pathogenesis of atherosclerosis: a perspective for the 1990s. *Nature, 362*(6423), 801-809. doi: 10.1038/362801a0

Ross, R. (1999). Atherosclerosis--an inflammatory disease. *N Engl J Med, 340*(2), 115-126. doi: 10.1056/NEJM199901143400207

Ross, R., & Glomset, J. A. (1973). Atherosclerosis and the arterial smooth muscle cell: Proliferation of smooth muscle is a key event in the genesis of the lesions of atherosclerosis. *Science, 180*(4093), 1332-1339.

Ross, R., & Glomset, J. A. (1976a). The pathogenesis of atherosclerosis (first of two parts). *N Engl J Med, 295*(7), 369-377. doi: 10.1056/NEJM197608122950707

Ross, R., & Glomset, J. A. (1976b). The pathogenesis of atherosclerosis (second of two parts). *N Engl J Med, 295*(8), 420-425. doi: 10.1056/NEJM197608192950805

Roth, E., Marczin, N., Balatonyi, B., Ghosh, S., Kovacs, V., Alotti, N., . . . Gasz, B. (2011). Effect of a glutathione S-transferase inhibitor on oxidative stress and ischemia-reperfusion-induced apoptotic signalling of cultured cardiomyocytes. *Exp Clin Cardiol, 16*(3), 92-96.

Rudic, R. D., McNamara, P., Curtis, A. M., Boston, R. C., Panda, S., Hogenesch, J. B., & Fitzgerald, G. A. (2004). BMAL1 and CLOCK, two essential components of the circadian clock, are involved in glucose homeostasis. *PLoS.Biol., 2*(11), e377. doi: 10.1371/journal.pbio.0020377 [doi]

Sabeti, P. C., Schaffner, S. F., Fry, B., Lohmueller, J., Varilly, P., Shamovsky, O., . . . Lander, E. S. (2006). Positive natural selection in the human lineage. *Science, 312*(5780), 1614-1620. doi: 312/5780/1614 [pii];10.1126/science.1124309 [doi]

Sanna, S., Jackson, A. U., Nagaraja, R., Willer, C. J., Chen, W. M., Bonnycastle, L. L., . . . Mohlke, K. L. (2008). Common variants in the GDF5-UQCC region are associated with variation in human height. *Nat Genet, 40*(2), 198-203. doi: 10.1038/ng.74

Scheer, F. A., Hilton, M. F., Mantzoros, C. S., & Shea, S. A. (2009). Adverse metabolic and cardiovascular consequences of circadian misalignment. *Proc.Natl.Acad.Sci.U.S.A, 106*(11), 4453-4458. doi: 0808180106 [pii];10.1073/pnas.0808180106 [doi]

Scherf, U., Ross, D. T., Waltham, M., Smith, L. H., Lee, J. K., Tanabe, L., . . . Weinstein, J. N. (2000). A gene expression database for the molecular pharmacology of cancer. *Nat Genet, 24*(3), 236-244. doi: 10.1038/73439

Schnabel, R. B., Lunetta, K. L., Larson, M. G., Dupuis, J., Lipinska, I., Rong, J., . . . Benjamin, E. J. (2009). The relation of genetic and environmental factors to systemic inflammatory biomarker concentrations. *Circ.Cardiovasc.Genet., 2*(3), 229-237. doi: CIRCGENETICS.108.804245 [pii];10.1161/CIRCGENETICS.108.804245 [doi]

Selhub, J. (1999). Homocysteine metabolism. *Annu Rev Nutr, 19*, 217-246. doi: 10.1146/annurev.nutr.19.1.217

Shaikh, T. H., Gai, X., Perin, J. C., Glessner, J. T., Xie, H., Murphy, K., . . . Hakonarson, H. (2009). High-resolution mapping and analysis of copy number variations in the human genome: a data resource for clinical and research applications. *Genome Res, 19*(9), 1682-1690. doi: 10.1101/gr.083501.108

Sharp, A. J., Locke, D. P., McGrath, S. D., Cheng, Z., Bailey, J. A., Vallente, R. U., . . . Eichler, E. E. (2005). Segmental duplications and copy-number variation in the human genome. *Am J Hum Genet, 77*(1), 78-88. doi: 10.1086/431652

Sheehan, N. A., Didelez, V., Burton, P. R., & Tobin, M. D. (2008). Mendelian randomisation and causal inference in observational epidemiology. *PLoS Med, 5*(8), e177. doi: 10.1371/journal.pmed.0050177

Sheehan, N. A., Meng, S., & Didelez, V. (2011). Mendelian randomisation: a tool for assessing causality in observational epidemiology. *Methods Mol Biol, 713*, 153-166. doi: 10.1007/978-1-60327-416-6_12

Shen, Y., Rattan, V., Sultana, C., & Kalra, V. K. (1996). Cigarette smoke condensate-induced adhesion molecule expression and transendothelial migration of monocytes. *Am J Physiol, 270*(5 Pt 2), H1624-1633.

Shenouda, S. M., & Vita, J. A. (2007). Effects of flavonoid-containing beverages and EGCG on endothelial function. *J Am Coll Nutr, 26*(4), 366S-372S.

Shimba, S., Ishii, N., Ohta, Y., Ohno, T., Watabe, Y., Hayashi, M., . . . Tezuka, M. (2005). Brain and muscle Arnt-like protein-1 (BMAL1), a component of the molecular clock, regulates adipogenesis. *Proc.Natl.Acad.Sci.U.S.A, 102*(34), 12071-12076. doi: 0502383102 [pii];10.1073/pnas.0502383102 [doi]

Shriner, D., Adeyemo, A., Gerry, N. P., Herbert, A., Chen, G., Doumatey, A., . . . Rotimi, C. N. (2009). Transferability and fine-mapping of genome-wide associated loci for adult height across human populations. *PLoS.One., 4*(12), e8398. doi: 10.1371/journal.pone.0008398 [doi]

Silverstein, R. L. (2009). Inflammation, atherosclerosis, and arterial thrombosis: role of the scavenger receptor CD36. *Cleve.Clin.J.Med., 76 Suppl 2*, S27-S30. doi: 76/Suppl_2/S27 [pii];10.3949/ccjm.76.s2.06 [doi]

Sing, C. F., Stengard, J. H., & Kardia, S. L. (2003). Genes, environment, and cardiovascular disease. *Arterioscler Thromb Vasc Biol, 23*(7), 1190-1196. doi: 10.1161/01.ATV.0000075081.51227.86

Slade, G. D., Offenbacher, S., Beck, J. D., Heiss, G., & Pankow, J. S. (2000). Acute-phase inflammatory response to periodontal disease in the US population. *J.Dent.Res., 79*(1), 49-57.

Smith, C. J., & Fischer, T. H. (2001). Particulate and vapor phase constituents of cigarette mainstream smoke and risk of myocardial infarction. *Atherosclerosis, 158*(2), 257-267.

Smith, C. M., Kelsey, K. T., Wiencke, J. K., Leyden, K., Levin, S., & Christiani, D. C. (1994). Inherited glutathione-S-transferase deficiency is a risk factor for pulmonary asbestosis. *Cancer Epidemiol Biomarkers Prev, 3*(6), 471-477.

Smulders, Y. M., de Man, A. M., Stehouwer, C. D., & Slaats, E. H. (1998). Trimethoprim and fasting plasma homocysteine. *Lancet, 352*(9143), 1827-1828. doi: 10.1016/S0140-6736(05)79890-1

Sobrin, L., Green, T., Sim, X., Jensen, R. A., Tai, E. S., Tay, W. T., . . . Wong, T. Y. (2011). Candidate gene association study for diabetic retinopathy in persons with type 2 diabetes: the Candidate gene Association Resource (CARe). *Invest Ophthalmol Vis Sci, 52*(10), 7593-7602. doi: 10.1167/iovs.11-7510

Sorescu, D., Weiss, D., Lassegue, B., Clempus, R. E., Szocs, K., Sorescu, G. P., . . . Griendling, K. K. (2002). Superoxide production and expression of nox family proteins in human atherosclerosis. *Circulation, 105*(12), 1429-1435.

Stamler, J. S., Osborne, J. A., Jaraki, O., Rabbani, L. E., Mullins, M., Singel, D., & Loscalzo, J. (1993). Adverse vascular effects of homocysteine are modulated by endothelium-derived relaxing factor and related oxides of nitrogen. *J Clin Invest, 91*(1), 308-318. doi: 10.1172/JCI116187

Starkebaum, G., & Harlan, J. M. (1986). Endothelial cell injury due to copper-catalyzed hydrogen peroxide generation from homocysteine. *J Clin Invest, 77*(4), 1370-1376. doi: 10.1172/JCI112442

Steinberg, D. (1997). Low density lipoprotein oxidation and its pathobiological significance. *J Biol Chem, 272*(34), 20963-20966.

Strange, R. C., Jones, P. W., & Fryer, A. A. (2000). Glutathione S-transferase: genetics and role in toxicology. *Toxicol Lett, 112-113*, 357-363.

Strange, R. C., Lear, J. T., & Fryer, A. A. (1998). Glutathione S-transferase polymorphisms: influence on susceptibility to cancer. *Chem Biol Interact, 111-112*, 351-364.

Studies, N.-N. W. G. o. R. i. A., Chanock, S. J., Manolio, T., Boehnke, M., Boerwinkle, E., Hunter, D. J., . . . Collins, F. S. (2007). Replicating genotype-phenotype associations. *Nature, 447*(7145), 655-660. doi: 10.1038/447655a

Subrahmanyan, L., Eberle, M. A., Clark, A. G., Kruglyak, L., & Nickerson, D. A. (2001). Sequence variation and linkage disequilibrium in the human T-cell receptor beta (TCRB) locus. *Am J Hum Genet, 69*(2), 381-395. doi: 10.1086/321297

Sun, Y., Scavini, M., Orlando, R. A., Murata, G. H., Servilla, K. S., Tzamaloukas, A. H., . . . Zager, P. G. (2010). Increased CD36 expression signals monocyte activation among patients with type 2 diabetes. *Diabetes Care, 33*(9), 2065-2067. doi: dc10-0460 [pii];10.2337/dc10-0460 [doi]

Tabor, H. K., Risch, N. J., & Myers, R. M. (2002). Candidate-gene approaches for studying complex genetic traits: practical considerations. *Nat Rev Genet, 3*(5), 391-397. doi: 10.1038/nrg796

Tamer, L., Ercan, B., Camsari, A., Yildirim, H., Cicek, D., Sucu, N., . . . Atik, U. (2004). Glutathione S-transferase gene polymorphism as a susceptibility factor in smoking-related coronary artery disease. *Basic Res Cardiol, 99*(3), 223-229. doi: 10.1007/s00395-004-0465-8

Tanaka, T., Nakata, T., Oka, T., Ogawa, T., Okamoto, F., Kusaka, Y., . . . Itakura, K. (2001). Defect in human myocardial long-chain fatty acid uptake is caused by FAT/CD36 mutations. *J.Lipid Res., 42*(5), 751-759
.
Tanaka, T., Scheet, P., Giusti, B., Bandinelli, S., Piras, M. G., Usala, G., . . . Ferrucci, L. (2009). Genome-wide association study of vitamin B6, vitamin B12, folate, and homocysteine blood concentrations. *Am J Hum Genet, 84*(4), 477-482. doi: 10.1016/j.ajhg.2009.02.011

Tappia, P. S., Troughton, K. L., Langley-Evans, S. C., & Grimble, R. F. (1995). Cigarette smoking influences cytokine production and antioxidant defences. *Clin Sci (Lond), 88*(4), 485-489.

Tawakol, A., Omland, T., Gerhard, M., Wu, J. T., & Creager, M. A. (1997). Hyperhomocyst(e)inemia is associated with impaired endothelium-dependent vasodilation in humans. *Circulation, 95*(5), 1119-1121
.
Terwilliger, J. D., & Goring, H. H. (2000). Gene mapping in the 20th and 21st centuries: statistical methods, data analysis, and experimental design. *Hum Biol, 72*(1), 63-132.

Thompson, S. G., Kienast, J., Pyke, S. D., Haverkate, F., & van de Loo, J. C. (1995). Hemostatic factors and the risk of myocardial infarction or sudden death in patients with angina pectoris. European Concerted Action on Thrombosis and Disabilities Angina Pectoris Study Group. *N Engl J Med, 332*(10), 635-641. doi: 10.1056/NEJM199503093321003

Thorand, B., Kolb, H., Baumert, J., Koenig, W., Chambless, L., Meisinger, C., . . . Herder, C. (2005). Elevated levels of interleukin-18 predict the development of type 2 diabetes: results from the MONICA/KORA Augsburg Study, 1984-2002. *Diabetes, 54*(10), 2932-2938. doi: 54/10/2932 [pii

Tian, C., Gregersen, P. K., & Seldin, M. F. (2008). Accounting for ancestry: population substructure and genome-wide association studies. *Hum Mol Genet, 17*(R2), R143-150. doi: 10.1093/hmg/ddn268

Tian, C., Plenge, R. M., Ransom, M., Lee, A., Villoslada, P., Selmi, C., . . . Seldin, M. F. (2008). Analysis and application of European genetic substructure using 300 K SNP information. *PLoS Genet, 4*(1), e4. doi: 10.1371/journal.pgen.0040004

Tirona, R. G., & Pang, K. S. (1999). Bimolecular glutathione conjugation kinetics of ethacrynic acid in rat liver: in vitro and perfusion studies. *J Pharmacol Exp Ther, 290*(3), 1230-1241.

Townsend, D. M., Findlay, V. L., & Tew, K. D. (2005). Glutathione S-transferases as regulators of kinase pathways and anticancer drug targets. *Methods Enzymol, 401*, 287-307. doi: 10.1016/S0076-6879(05)01019-0

Townsend, D. M., Tew, K. D., & Tapiero, H. (2003). The importance of glutathione in human disease. *Biomed Pharmacother, 57*(3-4), 145-155.

Tracy, R. P., Psaty, B. M., Macy, E., Bovill, E. G., Cushman, M., Cornell, E. S., & Kuller, L. H. (1997). Lifetime smoking exposure affects the association of C-reactive protein with cardiovascular disease risk factors and subclinical disease in healthy elderly subjects. *Arterioscler.Thromb.Vasc.Biol., 17*(10), 2167-2176.

Tracy, R. P., Psaty, B. M., Macy, E., Bovill, E. G., Cushman, M., Cornell, E. S., & Kuller, L. H. (1997). Lifetime smoking exposure affects the association of C-reactive protein with cardiovascular disease risk factors and subclinical disease in healthy elderly subjects. *Arterioscler Thromb Vasc Biol, 17*(10), 2167-2176.

Tsai, M. Y., Loria, C. M., Cao, J., Kim, Y., Siscovick, D. S., Schreiner, P. J., & Hanson, N. Q. (2009). Polygenic association with total homocysteine in the post-folic acid fortification era: the CARDIA study. *Mol Genet Metab, 98*(1-2), 181-186. doi: 10.1016/j.ymgme.2009.05.012

Turkanoglu, A., Can Demirdogen, B., Demirkaya, S., Bek, S., & Adali, O. (2010). Association analysis of GSTT1, GSTM1 genotype polymorphisms and serum total GST activity with ischemic stroke risk. *Neurol Sci, 31*(6), 727-734. doi: 10.1007/s10072-010-0330-5

Uchimoto, S., Tsumura, K., Hayashi, T., Suematsu, C., Endo, G., Fujii, S., & Okada, K. (1999). Impact of cigarette smoking on the incidence of Type 2 diabetes mellitus in middle-aged Japanese men: the Osaka Health Survey. *Diabet Med, 16*(11), 951-955.

Unoki, H., Takahashi, A., Kawaguchi, T., Hara, K., Horikoshi, M., Andersen, G., . . . Maeda, S. (2008). SNPs in KCNQ1 are associated with susceptibility to type 2 diabetes in East Asian and European populations. *Nat.Genet., 40*(9), 1098-1102. doi: ng.208 [pii];10.1038/ng.208 [doi]

Unverdorben, M., von Holt, K., & Winkelmann, B. R. (2009). Smoking and atherosclerotic cardiovascular disease: part II: role of cigarette smoking in cardiovascular disease development. *Biomark Med, 3*(5), 617-653. doi: 10.2217/bmm.09.51

van der Wal, A. C., Das, P. K., Bentz van de Berg, D., van der Loos, C. M., & Becker, A. E. (1989). Atherosclerotic lesions in humans. In situ immunophenotypic analysis suggesting an immune mediated response. *Lab Invest, 61*(2), 166-170.

van Leeuwen, W. M., Lehto, M., Karisola, P., Lindholm, H., Luukkonen, R., Sallinen, M., . . . Alenius, H. (2009). Sleep restriction increases the risk of developing cardiovascular diseases by augmenting proinflammatory responses through IL-17 and CRP. *PLoS.One., 4*(2), e4589. doi: 10.1371/journal.pone.0004589 [doi]

Vasan, R. S. (2006). Biomarkers of cardiovascular disease: molecular basis and practical considerations. *Circulation, 113*(19), 2335-2362. doi: 10.1161/CIRCULATIONAHA.104.482570

Vaughn, J. D., Bailey, L. B., Shelnutt, K. P., Dunwoody, K. M., Maneval, D. R., Davis, S. R., . . . Kauwell, G. P. (2004). Methionine synthase reductase 66A->G polymorphism is associated with increased plasma homocysteine concentration when combined with the homozygous methylenetetrahydrofolate reductase 677C->T variant. *J Nutr, 134*(11), 2985-2990.

Volanakis, J. E. (2001). Human C-reactive protein: expression, structure, and function. *Mol Immunol, 38*(2-3), 189-197.

Wald, D. S., Law, M., & Morris, J. K. (2002). Homocysteine and cardiovascular disease: evidence on causality from a meta-analysis. *BMJ, 325*(7374), 1202.

Wang, J., Zou, L., Huang, S., Lu, F., Lang, X., Han, L., . . . Xu, Z. (2010). Genetic polymorphisms of glutathione S-transferase genes GSTM1, GSTT1 and risk of coronary heart disease. *Mutagenesis, 25*(4), 365-369. doi: 10.1093/mutage/geq014

Wang, L., Jhee, K. H., Hua, X., DiBello, P. M., Jacobsen, D. W., & Kruger, W. D. (2004). Modulation of cystathionine beta-synthase level regulates total serum homocysteine in mice. *Circ Res, 94*(10), 1318-1324. doi: 10.1161/01.RES.0000129182.46440.4a

Wang, Q. (2005). Molecular genetics of coronary artery disease. *Curr Opin Cardiol, 20*(3), 182-188.

Wannamethee, S. G., Lowe, G. D., Shaper, A. G., Rumley, A., Lennon, L., & Whincup, P. H. (2005). Associations between cigarette smoking, pipe/cigar smoking, and smoking cessation, and haemostatic and inflammatory markers for cardiovascular disease. *Eur Heart J, 26*(17), 1765-1773. doi: 10.1093/eurheartj/ehi183

Weber, C., Erl, W., Weber, K., & Weber, P. C. (1996). Increased adhesiveness of isolated monocytes to endothelium is prevented by vitamin C intake in smokers. *Circulation, 93*(8), 1488-1492.

Welch, G. N., & Loscalzo, J. (1998). Homocysteine and atherothrombosis. *N Engl J Med, 338*(15), 1042-1050. doi: 10.1056/NEJM199804093381507

Welsh, J. B., Zarrinkar, P. P., Sapinoso, L. M., Kern, S. G., Behling, C. A., Monk, B. J., . . . Hampton, G. M. (2001). Analysis of gene expression profiles in normal and neoplastic ovarian tissue samples identifies candidate molecular markers of epithelial ovarian cancer. *Proc Natl Acad Sci U S A, 98*(3), 1176-1181. doi: 10.1073/pnas.98.3.1176

Wensley, F., Gao, P., Burgess, S., Kaptoge, S., Di, A. E., Shah, T., . . . Danesh, J. (2011). Association between C reactive protein and coronary heart disease: mendelian randomisation analysis based on individual participant data. *BMJ, 342*, d548.

Wesley, R. B., 2nd, Meng, X., Godin, D., & Galis, Z. S. (1998). Extracellular matrix modulates macrophage functions characteristic to atheroma: collagen type I enhances acquisition of resident macrophage traits by human peripheral blood monocytes in vitro. *Arterioscler Thromb Vasc Biol, 18*(3), 432-440.

Whincup, P. H., Refsum, H., Perry, I. J., Morris, R., Walker, M., Lennon, L., . . . Ebrahim, S. B. (1999). Serum total homocysteine and coronary heart disease: prospective study in middle aged men. *Heart, 82*(4), 448-454.

White, D. L., Li, D., Nurgalieva, Z., & El-Serag, H. B. (2008). Genetic variants of glutathione S-transferase as possible risk factors for hepatocellular carcinoma: a HuGE systematic review and meta-analysis. *Am J Epidemiol, 167*(4), 377-389. doi: 10.1093/aje/kwm315

Wilcken, B., Bamforth, F., Li, Z., Zhu, H., Ritvanen, A., Renlund, M., . . . Botto, L. D. (2003). Geographical and ethnic variation of the 677C>T allele of 5,10 methylenetetrahydrofolate reductase (MTHFR): findings from over 7000 newborns from 16 areas world wide. *J Med Genet, 40*(8), 619-625
.

Willer, C. J., Li, Y., & Abecasis, G. R. (2010). METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics., 26*(17), 2190-2191. doi: btq340 [pii];10.1093/bioinformatics/btq340 [doi]

Willer, C. J., Sanna, S., Jackson, A. U., Scuteri, A., Bonnycastle, L. L., Clarke, R., . . . Abecasis, G. R. (2008). Newly identified loci that influence lipid concentrations and risk of coronary artery disease. *Nat Genet, 40*(2), 161-169. doi: 10.1038/ng.76

Williams, K. T., & Schalinske, K. L. (2010). Homocysteine metabolism and its relation to health and disease. *Biofactors, 36*(1), 19-24. doi: 10.1002/biof.71

Wilson, P. W., Nam, B. H., Pencina, M., D'Agostino, R. B., Sr., Benjamin, E. J., & O'Donnell, C. J. (2005). C-reactive protein and risk of cardiovascular disease in men and women from the Framingham Heart Study. *Arch Intern Med, 165*(21), 2473-2478. doi: 10.1001/archinte.165.21.2473

Wingender, E., Chen, X., Hehl, R., Karas, H., Liebich, I., Matys, V., . . . Schacherer, F. (2000). TRANSFAC: an integrated system for gene expression regulation. *Nucleic Acids Res, 28*(1), 316-319.

Woo, K. S., Chook, P., Lolin, Y. I., Cheung, A. S., Chan, L. T., Sun, Y. Y., . . . Celermajer, D. S. (1997). Hyperhomocyst(e)inemia is a risk factor for arterial endothelial dysfunction in humans. *Circulation, 96*(8), 2542-2544.

Wu, G., Fang, Y. Z., Yang, S., Lupton, J. R., & Turner, N. D. (2004). Glutathione metabolism and its implications for health. *J Nutr, 134*(3), 489-492.

Xu, Y., & Whitmer, K. (2006). C-reactive protein and cardiovascular disease in people with diabetes: high-sensitivity CRP testing can help assess risk for future cardiovascular disease events in this population. *Am.J.Nurs., 106*(8), 66-72. doi: 00000446-200608000-00027 [pii]

Yang, Q., Bailey, L., Clarke, R., Flanders, W. D., Liu, T., Yesupriya, A., . . . Friedman, J. M. (2012). Prospective study of methylenetetrahydrofolate reductase (MTHFR) variant C677T and risk of all-cause and cardiovascular disease mortality among 6000 US adults. *Am J Clin Nutr, 95*(5), 1245-1253. doi: 10.3945/ajcn.111.022384

Yasuda, K., Miyake, K., Horikawa, Y., Hara, K., Osawa, H., Furuta, H., . . . Kasuga, M. (2008). Variants in KCNQ1 are associated with susceptibility to type 2 diabetes mellitus. *Nat.Genet., 40*(9), 1092-1097. doi: ng.207 [pii];10.1038/ng.207 [doi]

Ye, Z., Song, H., Higgins, J. P., Pharoah, P., & Danesh, J. (2006). Five glutathione s-transferase gene variants in 23,452 cases of lung cancer and 30,397 controls: meta-analysis of 130 studies. *PLoS Med, 3*(4), e91. doi: 10.1371/journal.pmed.0030091

Yeh, E. T., & Willerson, J. T. (2003). Coming of age of C-reactive protein: using inflammation markers in cardiology. *Circulation, 107*(3), 370-371.

Yusuf, S., Hawken, S., Ounpuu, S., Dans, T., Avezum, A., Lanas, F., . . . Investigators, I. S. (2004). Effect of potentially modifiable risk factors associated with myocardial infarction in 52 countries (the INTERHEART study): case-control study. *Lancet, 364*(9438), 937-952. doi: 10.1016/S0140-6736(04)17018-9

Zhao, Y., Marotta, M., Eichler, E. E., Eng, C., & Tanaka, H. (2009). Linkage disequilibrium between two high-frequency deletion polymorphisms: implications for association studies involving the glutathione-S transferase (GST) genes. *PLoS Genet, 5*(5), e1000472. doi: 10.1371/journal.pgen.1000472